

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/40, C07K 14/18, C12Q 1/70,</b> <b>C07K 16/10, A61K 39/29</b>		<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 99/04008</b> <b>(43) International Publication Date:</b> 28 January 1999 (28.01.99)
<b>(21) International Application Number:</b> PCT/US98/14688 <b>(22) International Filing Date:</b> 16 July 1998 (16.07.98)  <b>(30) Priority Data:</b> 60/053,062 18 July 1997 (18.07.97) US 09/014,416 27 January 1998 (27.01.98) US  <b>(71) Applicant:</b> THE GOVERNMENT OF THE UNITED STATES OF AMERICA, asrepresented by THE SECRETARY, DEPARTMENT OF HEALTHAND HUMAN SERVICES [US/US]; Office of Technology Transfer, National Institutes of Health, Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).  <b>(72) Inventors:</b> YANAGI, Masayuki; 257 Congressional Lane, #402, Rockville, MD 20852 (US). BUKH, Jens; 7 Center Drive, MSC 0740, Bethesda, MD 20892 (US). EMERSON, Suzanne, U.; 18201 Woodcrest Drive, Rockville, MD 20852 (US). PURCELL, Robert, H.; 17517 White Grounds Road, Boyds, MD 20841 (US).  <b>(74) Agents:</b> FEILER, William, S. et al.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	
<b>(54) Title:</b> CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND USES THEREOF			
<b>(57) Abstract</b>  The present invention discloses nucleic acid sequences which encode infectious hepatitis C viruses and the use of these sequences, and polypeptides encoded by all or part of these sequences, in the development of vaccines and diagnostics for HCV and in the development of screening assays for the identification of antiviral agents for HCV.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

Title Of Invention

Cloned Genomes Of Infectious  
Hepatitis C Viruses And Uses Thereof

5

This application claims the benefit of U.S.  
Provisional Application No. 60/053,062 filed July 18,  
1997.

Field Of Invention

10

The present invention relates to molecular  
approaches to the production of nucleic acid sequences  
which comprise the genome of infectious hepatitis C  
viruses. In particular, the invention provides nucleic  
acid sequences which comprise the genomes of infectious  
hepatitis C viruses of genotype 1a and 1b strains. The  
invention therefore relates to the use of these sequences,  
and polypeptides encoded by all or part of these  
sequences, in the development of vaccines and diagnostic  
assays for HCV and in the development of screening assays  
for the identification of antiviral agents for HCV.

15

20

Background Of Invention

25

Hepatitis C virus (HCV) has a positive-sense  
single-strand RNA genome and is a member of the virus  
family *Flaviviridae* (Choo et al., 1991; Rice, 1996). As  
for all positive-stranded RNA viruses, the genome of HCV  
functions as mRNA from which all viral proteins necessary  
for propagation are translated.

30

The viral genome of HCV is approximately 9600  
nucleotides (nts) and consists of a highly conserved 5'  
untranslated region (UTR), a single long open reading

35

- 2 -

frame (ORF) of approximately 9,000 nts and a complex 3' UTR. The 5' UTR contains an internal ribosomal entry site (Tsukiyama-Kohara et al., 1992; Honda et al., 1996). The 3' UTR consists of a short variable region, a polypyrimidine tract of variable length and, at the 3' end, a highly conserved region of approximately 100 nts (Kolykhalov et al., 1996; Tanaka et al., 1995; Tanaka et al., 1996; Yamada et al., 1996). The last 46 nucleotides of this conserved region were predicted to form a stable stem-loop structure thought to be critical for viral replication (Blight and Rice, 1997; Ito and Lai, 1997; Tsuchihara et al., 1997). The ORF encodes a large polypeptide precursor that is cleaved into at least 10 proteins by host and viral proteinases (Rice, 1996). The predicted envelope proteins contain several conserved N-linked glycosylation sites and cysteine residues (Okamoto et al., 1992a). The NS3 gene encodes a serine protease and an RNA helicase and the NS5B gene encodes an RNA-dependent RNA polymerase.

Globally, six major HCV genotypes (genotypes 1-6) and multiple subtypes (a, b, c, etc.) have been identified (Bukh et al., 1993; Simmonds et al., 1993). The most divergent HCV isolates differ from each other by more than 30% over the entire genome (Okamoto et al., 1992a) and HCV circulates in an infected individual as a quasispecies of closely related genomes (Bukh et al., 1995; Farci et al., 1997).

At present, more than 80% of individuals infected with HCV become chronically infected and these chronically infected individuals have a relatively high

35

- 3 -

risk of developing chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (Hoofnagle, 1997). In the U.S., HCV genotypes 1a and 1b constitute the majority of infections while in many other areas, especially in Europe and Japan, genotype 1b predominates.

The only effective therapy for chronic hepatitis C, interferon (IFN), induces a sustained response in less than 25% of treated patients (Fried and Hoofnagle, 1995). Consequently, HCV is currently the most common cause of end stage liver failure and the reason for about 30% of liver transplants performed in the U.S. (Hoofnagle, 1997). In addition, a number of recent studies suggested that the severity of liver disease and the outcome of therapy may be genotype-dependent (reviewed in Bukh et al., 1997). In particular, these studies suggested that infection with HCV genotype 1b was associated with more severe liver disease (Brechot, 1997) and a poorer response to IFN therapy (Fried and Hoofnagle, 1995). As a result of the inability to develop a universally effective therapy against HCV infection, it is estimated that there are still more than 25,000 new infections yearly in the U.S. (Alter 1997) Moreover, since there is no vaccine for HCV, HCV remains a serious public health problem.

However, despite the intense interest in the development of vaccines and therapies for HCV, progress has been hindered by the absence of a useful cell culture system and the lack of any small animal model for laboratory study. For example, while replication of HCV in several cell lines has been reported, such observations have turned out not to be highly reproducible. In

- 4 -

addition, the chimpanzee is the only animal model, other than man, for this disease. Consequently, HCV has been able to be studied only by using clinical materials obtained from patients or experimentally infected chimpanzees (an animal model whose availability is very limited).

However, several researchers have recently reported the construction of infectious cDNA clones of HCV, the identification of which would permit a more effective search for susceptible cell lines and facilitate molecular analysis of the viral genes and their function. For example, Dash et al., (1997) and Yoo et al., (1995) reported that RNA transcripts from cDNA clones of HCV-1 (genotype 1a) and HCV-N (genotype 1b), respectively, resulted in viral replication after transfection into human hepatoma cell lines. Unfortunately, the viability of these clones was not tested in vivo and concerns were raised about the infectivity of these cDNA clones in vitro (Fausto, 1997). In addition, both clones did not contain the terminal 98 conserved nucleotides at the very 3' end of the UTR.

Kolykhalov et al., (1997) and Yanagi et al. (1997) reported the derivation from HCV strain H77 (which is genotype 1a) of cDNA clones of HCV that are infectious for chimpanzees. However, while these infectious clones will aid in studying HCV replication and pathogenesis and will provide an important tool for development of in vitro replication and propagation systems, it is important to have infectious clones of more than one genotype given the extensive genetic heterogeneity of HCV and the potential

35

- 5 -

impact of such heterogeneity on the development of effective therapies and vaccines for HCV.

### Summary Of The Invention

5           The present invention relates to nucleic acid sequences which comprise the genome of infectious hepatitis C viruses and in particular, nucleic acid sequences which comprise the genome of infectious  
10 hepatitis C viruses of genotype 1a and 1b strains. It is therefore an object of the invention to provide nucleic acid sequences which encode infectious hepatitis C viruses. Such nucleic acid sequences are referred to  
15 throughout the application as "infectious nucleic acid sequences".

          For the purposes of this application, nucleic acid sequence refers to RNA, DNA, cDNA or any variant thereof capable of directing host organism synthesis of a  
20 hepatitis C virus polypeptide. It is understood that nucleic acid sequence encompasses nucleic acid sequences, which due to degeneracy, encode the same polypeptide sequence as the nucleic acid sequences described herein.

25           The invention also relates to the use of the infectious nucleic acid sequences to produce chimeric genomes consisting of portions of the open reading frames of infectious nucleic acid sequences of other genotypes (including, but not limited to, genotypes 1, 2, 3, 4, 5  
30 and 6) and subtypes (including, but not limited to, subtypes 1a, 1b, 2a, 2b, 2c, 3a 4a-4f, 5a and 6a) of HCV. For example infectious nucleic acid sequence of the 1a and 1b strains H77 and HC-J4, respectively, described herein

35

- 6 -

can be used to produce chimeras with sequences from the genomes of other strains of HCV from different genotypes or subtypes. Nucleic acid sequences which comprise sequence from the open-reading frames of 2 or more HCV genotypes or subtypes are designated "chimeric nucleic acid sequences".

The invention further relates to mutations of the infectious nucleic acid sequences of the invention where mutation includes, but is not limited to, point mutations, deletions and insertions. In one embodiment, a gene or fragment thereof can be deleted to determine the effect of the deleted gene or genes on the properties of the encoded virus such as its virulence and its ability to replicate. In an alternative embodiment, a mutation may be introduced into the infectious nucleic acid sequences to examine the effect of the mutation on the properties of the virus in the host cell.

The invention also relates to the introduction of mutations or deletions into the infectious nucleic acid sequences in order to produce an attenuated hepatitis C virus suitable for vaccine development.

The invention further relates to the use of the infectious nucleic acid sequences to produce attenuated viruses via passage in vitro or in vivo of the viruses produced by transfection of a host cell with the infectious nucleic acid sequence.

The present invention also relates to the use of the nucleic acid sequences of the invention or fragments thereof in the production of polypeptides where "nucleic acid sequences of the invention" refers to infectious



- 7 -

°  
nucleic acid sequences, mutations of infectious nucleic acid sequences, chimeric nucleic acid sequences and sequences which comprise the genome of attenuated viruses produced from the infectious nucleic acid sequences of the invention. The polypeptides of the invention, especially structural polypeptides, can serve as immunogens in the development of vaccines or as antigens in the development of diagnostic assays for detecting the presence of HCV in biological samples.

The invention therefore also relates to vaccines for use in immunizing mammals especially humans against hepatitis C. In one embodiment, the vaccine comprises one or more polypeptides made from a nucleic acid sequence of the invention or fragment thereof. In a second embodiment, the vaccine comprises a hepatitis C virus produced by transfection of host cells with the nucleic acid sequences of the invention.

The present invention therefore relates to methods for preventing hepatitis C in a mammal. In one embodiment the method comprises administering to a mammal a polypeptide or polypeptides encoded by a nucleic acid sequence of the invention in an amount effective to induce protective immunity to hepatitis C. In another embodiment, the method of prevention comprises administering to a mammal a hepatitis C virus of the invention in an amount effective to induce protective immunity against hepatitis C.

In yet another embodiment, the method of protection comprises administering to a mammal a nucleic acid sequence of the invention or a fragment thereof in an

- 8 -

amount effective to induce protective immunity against hepatitis C.

The invention also relates to hepatitis C viruses produced by host cells transfected with the nucleic acid sequences of the present invention.

The invention therefore also provides pharmaceutical compositions comprising the nucleic acid sequences of the invention and/or their encoded hepatitis C viruses. The invention further provides pharmaceutical compositions comprising polypeptides encoded by the nucleic acid sequences of the invention or fragments thereof. The pharmaceutical compositions of the invention may be used prophylactically or therapeutically.

The invention also relates to antibodies to the hepatitis C viruses of the invention or their encoded polypeptides and to pharmaceutical compositions comprising these antibodies.

The present invention further relates to polypeptides encoded by the nucleic acid sequences of the invention fragments thereof. In one embodiment, said polypeptide or polypeptides are fully or partially purified from hepatitis C virus produced by cells transfected with nucleic acid sequence of the invention. In another embodiment, the polypeptide or polypeptides are produced recombinantly from a fragment of the nucleic acid sequences of the invention. In yet another embodiment, the polypeptides are chemically synthesized.

The invention also relates to the use of the nucleic acid sequences of the invention to identify cell

35

- 9 -

lines capable of supporting the replication of HCV in vitro.

The invention further relates to the use of the nucleic acid sequences of the invention or their encoded proteases (e.g. NS3 protease) to develop screening assays to identify antiviral agents for HCV.

#### Brief Description Of Figures

Figure 1 shows a strategy for constructing full-length cDNA clones of HCV strain H77. The long PCR products amplified with H1 and H9417R primers were cloned directly into pGEM-9zf(-) after digestion with Not I and Xba I (pH21<sub>I</sub> and pH50<sub>I</sub>). Next, the 3' UTR was cloned into both pH21<sub>I</sub> and pH50<sub>I</sub> after digestion with Afl II and Xba<sup>-</sup>I (pH21 and pH50). pH21 was tested for infectivity in a chimpanzee. To improve the efficiency of cloning, we constructed a cassette vector with consensus 5' and 3' termini of H77. This cassette vector (pCV) was obtained by cutting out the BamHI fragment (nts 1358- 7530 of the H77 genome) from pH50, followed by religation. Finally, the long PCR products of H77 amplified with primers H1 and H9417R (H product) or primers A1 and H9417R (A product) were cloned into pCV after digestion with Age I and Afl II or with Pin AI and Bfr I. The latter procedure yielded multiple complete cDNA clones of strain H77 of HCV.

Figure 2 shows the results of gel electrophoresis of long RT-PCR amplicons of the entire ORF of H77 and the transcription mixture of the infectious clone of H77. The complete ORF was amplified by long RT-PCR with the primers H1 or A1 and H9417R from 10<sup>5</sup> GE of

- 10 -

H77. A total of 10  $\mu$ g of the consensus chimeric clone (pCV-H77C) linearized with Xba I was transcribed in a 100  $\mu$ l reaction with T7 RNA polymerase. Five  $\mu$ l of the transcription mixture was analyzed by gel electrophoresis and the remainder of the mixture was injected into a chimpanzee. Lane 1, molecular weight marker ; lane 2, products amplified with primers H1 and H9417R; lane 3, products amplified with primers A1 and H9417R; lane 4, transcription mixture containing the RNA transcripts and linearized clone pCV-H77C (12.5 kb).

Figure 3 is a diagram of the genome organization of HCV strain H77 and the genetic heterogeneity of individual full-length clones compared with the consensus sequence of H77. Solid lines represent aa changes. Dashed lines represent silent mutations. A \* in pH21 represents a point mutation at nt 58 in the 5' UTR. In the ORF, the consensus chimeric clone pCV-H77C had 11 nt differences [at positions 1625 (C→T), 2709 (T→C), 3380 (A→G), 3710 (C→T), 3914 (G→A), 4463 (T→C), 5058 (C→T), 5834 (C→T), 6734 (T→C), 7154 (C→T), and 7202 (T→C)] and one aa change (F → L at aa 790) compared with the consensus sequence of H77. This clone was infectious. Clone pH21 and pCV-H11 had 19 nts (7 aa) and 64 nts (21 aa) differences respectively, compared with the consensus sequence of H77. These two clones were not infectious. A single point mutation in the 3' UTR at nucleotide 9406 (G→A) introduced to create an Afl II cleavage site is not shown.

- 11 -

Figures 4A-4F show the complete nucleotide sequence of a H77C clone produced according to the present invention and Figures 4G-4H show the amino acid sequence encoded by the H77C clone.

5 Figure 5 shows an agarose gel of long RT-PCR amplicons and transcription mixtures. Lanes 1 and 4: Molecular weight marker (*Lambda/HindIII* digest). Lanes 2 and 3: RT-PCR amplicons of the entire ORF of HC-J4. Lane 10 5: pCV-H77C transcription control (Yanagi et al., 1997). Lanes 6, 7, and 8: 1/40 of each transcription mixture of pCV-J4L2S, pCV-J4L4S and pCV-J4L6S, respectively, which was injected into the chimpanzee.

15 Figure 6 shows the strategy utilized for the construction of full-length cDNA clones of HCV strain HC-J4. The long PCR products were cloned as two separate fragments (L and S) into a cassette vector (pCV) with fixed 5' and 3' termini of HCV (Yanagi et al., 1997). 20 Full-length cDNA clones of HC-J4 were obtained by inserting the L fragment from three pCV-J4L clones into three identical pCV-J4S9 clones after digestion with *PinAI* (isoschizomer of *AgeI*) and *BfrI* (isoschizomer of *AflIII*). 25

Figure 7 shows amino acid positions with a quasispecies of HC-J4 in the acute phase plasma pool obtained from an experimentally infected chimpanzee. Cons-p9: consensus amino acid sequence deduced from 30 analysis of nine L fragments and nine S fragments (see Fig. 6). Cons-D: consensus sequence derived from direct sequencing of the PCR product. A, B, C: groups of similar viral species. Dot: amino acid identical to that in Cons-

35

- 12 -

o

p9. Capital letter: amino acid different from that in Cons-p9. Cons-F: composite consensus amino acid sequence combining Cons-p9 and Cons-D. Boxed amino acid: different from that in Cons-F. Shaded amino acid: different from that in all species A sequences. An \*: defective ORF due to a nucleotide deletion (clone L1, aa 1097) or insertion (clone L7, aa 2770). Diagonal lines: fragments used to construct the infectious clone.

Figure 8 shows comparisons (percent difference) of nucleotide (nts. 156 - 8935) and predicted amino acid sequences (aa 1 - 2864) of L clones (species A, B, and C, this study), HC-J4/91 (Okamoto et al., 1992b) and HC-J4/83 (Okamoto et al., 1992b). Differences among species A sequences and among species B sequences are shaded.

Figure 9 shows UPGMA ("unweighted pair group method with arithmetic mean") trees of HC-J4/91 (Okamoto et al., 1992b), HC-J4/83 (Okamoto et al., 1992b), two prototype strains of genotype 1b (HCV-J, Kato et al., 1990; HCV-BK, Takamizawa et al., 1991), and L clones (this study).

Figure 10 shows the alignment of the HVR1 and HVR2 amino acid sequences of the E2 sequences of nine L clones of HC-J4 (species A, B, and C) obtained from an early acute phase plasma pool of an experimentally infected chimpanzee compared with the sequences of eight clones (HC-J4/91-20 through HC-J4/91-27, Okamoto et al., 1992b) derived from the inoculum. Dot: an amino acid identical to that in the top line. Capital letters: amino acid different from that in the top line.

35

- 13 -

Figure 11 shows the alignment of the 5' UTR and the 3' UTR sequences of infectious clones of genotype 1a (pCV-H77C) and 1b (pCV-J4L6S). Top line: consensus sequence of the indicated strain. Dot: identity with consensus sequence. Capital letter: different from the consensus sequence. Dash: deletion. Underlined: *P*inAI and *B*frI cleavage site. Numbering corresponds to the HCV sequence of pCV-J4L6S.

Figure 12 shows a comparison of individual full-length cDNA clones of the ORF of HCV strain HC-J4 with the consensus sequence (see Fig. 7). Solid lines: amino acid changes. Dashed lines: silent mutations. Clone pCV-J4L6S was infectious in vivo whereas clones pCV-J4L2S and pCV-J4L4S were not.

Figure 13 shows biochemical (ALT levels) and PCR analyses of a chimpanzee following percutaneous intrahepatic transfection with RNA transcripts of the infectious clone of pCV-J4L2S, pCV-J4L4S and pCV-J4L6S. The ALT serum enzyme levels were measured in units per liter (u/l). For the PCR analysis, "HCV RNA" represented by an open rectangle indicates a serum sample that was negative for HCV after nested PCR; "HCV RNA" represented by a closed rectangle indicates that the serum sample was positive for HCV and HCV GE titer on the right-hand y-axis represents genome equivalents.

Figures 14A-14F show the nucleotide sequence of the infectious clone of genotype 1b strain HC-J4 and Figures 14G-14H show the amino acid sequence encoded by the HC-J4 clone.

- 14 -

Figure 15 shows the strategy for constructing a chimeric HCV clone designated pH77CV-J4 which contains the nonstructural region of the infectious clone of genotype 1a strain H77 and the structural region of the infectious clone of genotype 1b strain HC-J4.

Figures 16A-16F show the nucleotide sequence of the chimeric 1a/1b clone pH77CV-J4 of Figure 15 and Figures 16G-16H show the amino acid sequence encoded by the chimeric 1a/1b clone.

Figures 17A and 17B show the sequence of the 3' untranslated region remaining in various 3' deletion mutants of the 1a infectious clone pCV-H77C and the strategy utilized in constructing each 3' deletion mutant (Figures 17C-17G).

Of the seven deletion mutants shown, two (pCV-H77C(-98X) and (-42X)) have been constructed and tested for infectivity in chimpanzees (see Figures 17A and 17C) and the other six are to be constructed and tested for infectivity as described in Figures 17D-17G.

Figures 18A and 18B show biochemical (ALT levels), PCR (HCV RNA and HCV GE titer), serological (anti-HCV) and histopathological (Fig. 18B only) analyses of chimpanzees 1494 (Fig. 18A) and 1530 (Fig. 18B) following transfection with the infectious cDNA clone pCV-H77C.

The ALT serum enzyme levels were measured in units per ml (u/l). For the PCR analysis, "HCV RNA" represented by an open rectangle indicates a serum sample that was negative for HCV after nested PCR; "HCV RNA" represented by a closed rectangle indicates that the serum



- 15 -

sample was positive for HCV; and HCV GE titer on the right-hand y-axis represents genome equivalents.

The bar marked "anti-HCV" indicates samples that were positive for anti-HCV antibodies as determined by commercial assays. The histopathology scores in Figure 18B correspond to no histopathology (O), mild hepatitis (●) and moderate to severe hepatitis (●).

#### DESCRIPTION OF THE INVENTION

The present invention relates to nucleic acid sequences which comprise the genome of an infectious hepatitis C virus. More specifically, the invention relates to nucleic acid sequences which encode infectious hepatitis C viruses of genotype 1a and 1b strains. In one embodiment, the infectious nucleic acid sequence of the invention has the sequence shown in Figures 4A-4F of this application. In another embodiment, the infectious nucleic acid sequence has the sequence shown in Figures 14A-14F and is contained in a plasmid construct deposited with the American Type Culture Collection (ATCC) on January 26, 1998 and having ATCC accession number 209596.

The invention also relates to "chimeric nucleic acid sequences" where the chimeric nucleic acid sequences consist of open-reading frame sequences taken from infectious nucleic acid sequences of hepatitis C viruses of different genotypes or subtypes.

In one embodiment, the chimeric nucleic acid sequence consists of sequence from the genome of an HCV strain belonging to one genotype or subtype which encodes structural polypeptides and sequence of an HCV strain

- 16 -

°  
belonging to another genotype strain or subtype which  
encodes nonstructural polypeptides. Such chimeras can be  
produced by standard techniques of restriction digestion,  
PCR amplification and subcloning known to those of  
5 ordinary skill in the art.

In a preferred embodiment, the sequence encoding  
nonstructural polypeptides is from an infectious nucleic  
acid sequence encoding a genotype 1a strain where the  
10 construction of a chimeric 1a/1b nucleic acid sequence is  
described in Example 9 and the chimeric 1a/1b nucleic acid  
sequence is shown in Figures 16A-16F. It is believed that  
the construction of such chimeric nucleic acid sequences  
will be of importance in studying the growth and virulence  
15 properties of hepatitis C virus and in the production of  
hepatitis C viruses suitable to confer protection against  
multiple genotypes of HCV. For example, one might produce  
a "multivalent" vaccine by putting epitopes from several  
20 genotypes or subtypes into one clone. Alternatively one  
might replace just a single gene from an infectious  
sequence with the corresponding gene from the genomic  
sequence of a strain from another genotype or subtype or  
create a chimeric gene which contains portions of a gene  
25 from two genotypes or subtypes. Examples of genes which  
could be replaced or which could be made chimeric,  
include, but are not limited to, the E1, E2 and NS4 genes.

The invention further relates to mutations of  
30 the infectious nucleic acid sequences where "mutations"  
includes, but is not limited to, point mutations,  
deletions and insertions. Of course, one of ordinary  
skill in the art would recognize that the size of the

35

- 17 -

°  
insertions would be limited by the ability of the resultant nucleic acid sequence to be properly packaged within the virion. Such mutation could be produced by techniques known to those of skill in the art such as  
5 site-directed mutagenesis, fusion PCR, and restriction digestion followed by religation.

In one embodiment, mutagenesis might be undertaken to determine sequences that are important for  
10 viral properties such as replication or virulence. For example, one may introduce a mutation into the infectious nucleic acid sequence which eliminates the cleavage site between the NS4A and NS4B polypeptides to examine the effects on viral replication and processing of the  
15 polypeptide. Alternatively, one or more of the 3 amino acids encoded by the infectious 1b nucleic acid sequence shown in Figures 14A-14F which differ from the HC-J4 consensus sequence may be back mutated to the  
20 corresponding amino acid in the HC-J4 consensus sequence to determine the importance of these three amino acid changes to infectivity or virulence. In yet another embodiment, one or more of the amino acids from the  
25 noninfectious 1b clones pCV-J4L2S and pCV-J4L4S which differ from the consensus sequence may be introduced into the infectious 1b sequence shown in Figures 14A-14F.

In yet another example, one may delete all or part of a gene or of the 5' or 3' nontranslated region  
30 contained in an infectious nucleic acid sequence and then transfect a host cell (animal or cell culture) with the mutated sequence and measure viral replication in the host by methods known in the art such as RT-PCR. Preferred  
35

- 18 -

genes include, but are not limited to, the P7, NS4B and NS5A genes. Of course, those of ordinary skill in the art will understand that deletion of part of a gene, preferably the central portion of the gene, may be preferable to deletion of the entire gene in order to conserve the cleavage site boundaries which exist between proteins in the HCV polyprotein and which are necessary for proper processing of the polyprotein.

In the alternative, if the transfection is into a host animal such as a chimpanzee, one can monitor the virulence phenotype of the virus produced by transfection of the mutated infectious nucleic acid sequence by methods known in the art such as measurement of liver enzyme levels (alanine aminotransferase (ALT) or isocitrate dehydrogenase (ICD)) or by histopathology of liver biopsies. Thus, mutations of the infectious nucleic acid sequences may be useful in the production of attenuated HCV strains suitable for vaccine use.

The invention also relates to the use of the infectious nucleic acid sequences of the present invention to produce attenuated viral strains via passage in vitro or in vivo of the virus produced by transfection with the infectious nucleic acid sequences.

The present invention therefore relates to the use of the nucleic acid sequences of the invention to identify cell lines capable of supporting the replication of HCV.

In particular, it is contemplated that the mutations of the infectious nucleic acid sequences of the invention and the production of chimeric sequences as

- 19 -

discussed above may be useful in identifying sequences critical for cell culture adaptation of HCV and hence, may be useful in identifying cell lines capable of supporting HCV replication.

5 Transfection of tissue culture cells with the nucleic acid sequences of the invention may be done by methods of transfection known in the art such as electroporation, precipitation with DEAE-Dextran or  
10 calcium phosphate or liposomes.

In one such embodiment, the method comprises the growing of animal cells, especially human cells, in vitro and transfecting the cells with the nucleic acid of the invention, then determining if the cells show indicia of  
15 HCV infection. Such indicia include the detection of viral antigens in the cell, for example, by immunofluorescent procedures well known in the art; the detection of viral polypeptides by Western blotting using  
20 antibodies specific therefor; and the detection of newly transcribed viral RNA within the cells via methods such as RT-PCR. The presence of live, infectious virus particles following such tests may also be shown by injection of  
25 cell culture medium or cell lysates into healthy, susceptible animals, with subsequent exhibition of the symptoms of HCV infection.

Suitable cells or cell lines for culturing HCV include, but are not limited to, lymphocyte and hepatocyte  
30 cell lines known in the art.

Alternatively, primary hepatocytes can be cultured, and then infected with HCV; or, the hepatocyte cultures could be derived from the livers of infected  
35

- 20 -

chimpanzees. In addition, various immortalization methods known to those of ordinary skill in the art can be used to obtain cell-lines derived from hepatocyte cultures. For example, primary hepatocyte cultures may be fused to a variety of cells to maintain stability.

The present invention further relates to the in vitro and in vivo production of hepatitis C viruses from the nucleic acid sequences of the invention.

In one embodiment, the sequences of the invention can be inserted into an expression vector that functions in eukaryotic cells. Eukaryotic expression vectors suitable for producing high efficiency gene transfer in vivo are well known to those of ordinary skill in the art and include, but are not limited to, plasmids, vaccinia viruses, retroviruses, adenoviruses and adeno-associated viruses.

In another embodiment, the sequences contained in the recombinant expression vector can be transcribed in vitro by methods known to those of ordinary skill in the art in order to produce RNA transcripts which encode the hepatitis C viruses of the invention. The hepatitis C viruses of the invention may then be produced by transfecting cells by methods known to those of ordinary skill in the art with either the in vitro transcription mixture containing the RNA transcripts (see Example 4) or with the recombinant expression vectors containing the nucleic acid sequences described herein.

The present invention also relates to the construction of cassette vectors useful in the cloning of viral genomes wherein said vectors comprise a nucleic acid

- 21 -

sequence to be cloned, and said vector reading in the correct phase for the expression of the viral nucleic acid to be cloned. Such a cassette vector will, of course, also possess a promoter sequence, advantageously placed upstream of the sequence to be expressed. Cassette vectors according to the present invention are constructed according to the procedure described in Figure 1, for example, starting with plasmid pCV. Of course, the DNA to be inserted into said cassette vector can be derived from any virus, advantageously from HCV, and most advantageously from the H77 strain of HCV. The nucleic acid to be inserted according to the present invention can, for example, contain one or more open reading frames of the virus, for example, HCV. The cassette vectors of the present invention may also contain, optionally, one or more expressible marker genes for expression as an indication of successful transfection and expression of the nucleic acid sequences of the vector. To insure expression, the cassette vectors of the present invention will contain a promoter sequence for binding of the appropriate cellular RNA polymerase, which will depend on the cell into which the vector has been introduced. For example, if the host cell is a bacterial cell, then said promoter will be a bacterial promoter sequence to which the bacterial RNA polymerases will bind.

The hepatitis C viruses produced from the sequences of the invention may be purified or partially purified from the transfected cells by methods known to those of ordinary skill in the art. In a preferred embodiment, the viruses are partially purified prior to

35

- 22 -

their use as immunogens in the pharmaceutical compositions and vaccines of the present invention.

The present invention therefore relates to the use of the hepatitis C viruses produced from the nucleic acid sequences of the invention as immunogens in live or killed (e.g., formalin inactivated) vaccines to prevent hepatitis C in a mammal.

In an alternative embodiment, the immunogen of the present invention may be an infectious nucleic acid sequence, a chimeric nucleic acid sequence, or a mutated infectious nucleic acid sequence which encodes a hepatitis C virus. Where the sequence is a cDNA sequence, the cDNAs and their RNA transcripts may be used to transfect a mammal by direct injection into the liver tissue of the mammal as described in the Examples.

Alternatively, direct gene transfer may be accomplished via administration of a eukaryotic expression vector containing a nucleic acid sequence of the invention.

In yet another embodiment, the immunogen may be a polypeptide encoded by the nucleic acid sequences of the invention. The present invention therefore also relates to polypeptides produced from the nucleic acid sequences of the invention or fragments thereof. In one embodiment, polypeptides of the present invention can be recombinantly produced by synthesis from the nucleic acid sequences of the invention or isolated fragments thereof, and purified, or partially purified, from transfected cells using methods already known in the art. In an alternative embodiment, the polypeptides may be purified or partially



- 23 -

o

5

purified from viral particles produced via transfection of a host cell with the nucleic acid sequences of the invention. Such polypeptides might, for example, include either capsid or envelope polypeptides prepared from the sequences of the present invention.

10

15

20

When used as immunogens, the nucleic acid sequences of the invention, or the polypeptides or viruses produced therefrom, are preferably partially purified prior to use as immunogens in pharmaceutical compositions and vaccines of the present invention. When used as a vaccine, the sequences and the polypeptide and virus products thereof, can be administered alone or in a suitable diluent, including, but not limited to, water, saline, or some type of buffered medium. The vaccine according to the present invention may be administered to an animal, especially a mammal, and most especially a human, by a variety of routes, including, but not limited to, intradermally, intramuscularly, subcutaneously, or in any combination thereof.

25

30

35

Suitable amounts of material to administer for prophylactic and therapeutic purposes will vary depending on the route selected and the immunogen (nucleic acid, virus, polypeptide) administered. One skilled in the art will appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. The vaccines of the present invention may be administered once or periodically until a suitable titer of anti-HCV antibodies appear in the blood. For an immunogen consisting of a nucleic acid sequence, a suitable amount of nucleic acid

- 24 -

sequence to be used for prophylactic purposes might be expected to fall in the range of from about 100 µg to about 5 mg and most preferably in the range of from about 500 µg to about 2mg. For a polypeptide, a suitable amount to use for prophylactic purposes is preferably 100 ng to 100 µg and for a virus  $10^2$  to  $10^6$  infectious doses. Such administration will, of course, occur prior to any sign of HCV infection.

A vaccine of the present invention may be employed in such forms as capsules, liquid solutions, suspensions or elixirs for oral administration, or sterile liquid forms such as solutions or suspensions. Any inert carrier is preferably used, such as saline or phosphate-buffered saline, or any such carrier in which the HCV of the present invention can be suitably suspended. The vaccines may be in the form of single dose preparations or in multi-dose flasks which can be utilized for mass-vaccination programs of both animals and humans. For purposes of using the vaccines of the present invention reference is made to Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, Pa., Osol (Ed.) (1980); and New Trends and Developments in Vaccines, Voller et al. (Eds.), University Park Press, Baltimore, Md. (1978), both of which provide much useful information for preparing and using vaccines. Of course, the polypeptides of the present invention, when used as vaccines, can include, as part of the composition or emulsion, a suitable adjuvant, such as alum (or aluminum hydroxide) when humans are to be vaccinated, to further stimulate production of antibodies by immune cells. When nucleic acids or viruses are used

- 25 -

for vaccination purposes, other specific adjuvants such as CpG motifs (Krieg, A.K. et al. (1995) and (1996)), may prove useful.

When the nucleic acids, viruses and polypeptides of the present invention are used as vaccines or inocula, they will normally exist as physically discrete units suitable as a unitary dosage for animals, especially mammals, and most especially humans, wherein each unit will contain a predetermined quantity of active material calculated to produce the desired immunogenic effect in association with the required diluent. The dose of said vaccine or inoculum according to the present invention is administered at least once. In order to increase the antibody level, a second or booster dose may be administered at some time after the initial dose. The need for, and timing of, such booster dose will, of course, be determined within the sound judgment of the administrator of such vaccine or inoculum and according to sound principles well known in the art. For example, such booster dose could reasonably be expected to be advantageous at some time between about 2 weeks to about 6 months following the initial vaccination. Subsequent doses may be administered as indicated.

The nucleic acid sequences, viruses and polypeptides of the present invention can also be administered for purposes of therapy, where a mammal, especially a primate, and most especially a human, is already infected, as shown by well known diagnostic measures. When the nucleic acid sequences, viruses or polypeptides of the present invention are used for such

35

- 26 -

therapeutic purposes, much of the same criteria will apply as when it is used as a vaccine, except that inoculation will occur post-infection. Thus, when the nucleic acid sequences, viruses or polypeptides of the present invention are used as therapeutic agents in the treatment of infection, the therapeutic agent comprises a pharmaceutical composition containing a sufficient amount of said nucleic acid sequences, viruses or polypeptides so as to elicit a therapeutically effective response in the organism to be treated. Of course, the amount of pharmaceutical composition to be administered will, as for vaccines, vary depending on the immunogen contained therein (nucleic acid, polypeptide, virus) and on the route of administration.

The therapeutic agent according to the present invention can thus be administered by, subcutaneous, intramuscular or intradermal routes. One skilled in the art will certainly appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. Of course, the actual amounts will vary depending on the route of administration as well as the sex, age, and clinical status of the subject which, in the case of human patients, is to be determined with the sound judgment of the clinician.

The therapeutic agent of the present invention can be employed in such forms as capsules, liquid solutions, suspensions or elixirs, or sterile liquid forms such as solutions or suspensions. Any inert carrier is preferably used, such as saline, phosphate-buffered

- 27 -

o

saline, or any such carrier in which the HCV of the present invention can be suitably suspended. The therapeutic agents may be in the form of single dose preparations or in the multi-dose flasks which can be  
5 utilized for mass-treatment programs of both animals and humans. Of course, when the nucleic acid sequences, viruses or polypeptides of the present invention are used as therapeutic agents they may be administered as a single  
10 dose or as a series of doses, depending on the situation as determined by the person conducting the treatment.

The nucleic acids, polypeptides and viruses of the present invention can also be utilized in the production of antibodies against HCV. The term "antibody"  
15 is herein used to refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules. Examples of antibody molecules are intact immunoglobulin molecules, substantially intact  
20 immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the art as Fab, F(ab')<sub>2</sub> and F(v) as well as chimeric antibody molecules.

25 Thus, the polypeptides, viruses and nucleic acid sequences of the present invention can be used in the generation of antibodies that immunoreact (i.e., specific binding between an antigenic determinant-containing molecule and a molecule containing an antibody combining  
30 site such as a whole antibody molecule or an active portion thereof) with antigenic determinants on the surface of hepatitis C virus particles.

35

- 28 -

°

The present invention therefore also relates to antibodies produced following immunization with the nucleic acid sequences, viruses or polypeptides of the present invention. These antibodies are typically  
5 produced by immunizing a mammal with an immunogen or vaccine to induce antibody molecules having immunospecificity for polypeptides or viruses produced in response to infection with the nucleic acid sequences of the present invention. When used in generating such  
10 antibodies, the nucleic acid sequences, viruses, or polypeptides of the present invention may be linked to some type of carrier molecule. The resulting antibody molecules are then collected from said mammal. Antibodies  
15 produced according to the present invention have the unique advantage of being generated in response to authentic, functional polypeptides produced according to the actual cloned HCV genome.

20 The antibody molecules of the present invention may be polyclonal or monoclonal. Monoclonal antibodies are readily produced by methods well known in the art. Portions of immunoglobulin molecules, such as Fabs, as well as chimeric antibodies, may also be produced by methods  
25 well known to those of ordinary skill in the art of generating such antibodies.

The antibodies according to the present invention may also be contained in blood plasma, serum,  
30 hybridoma supernatants, and the like. Alternatively, the antibody of the present invention is isolated to the extent desired by well known techniques such as, for example, using DEAE Sephadex. The antibodies produced

35

- 29 -

according to the present invention may be further purified so as to obtain specific classes or subclasses of antibody such as IgM, IgG, IgA, and the like. Antibodies of the IgG class are preferred for purposes of passive protection.

The antibodies of the present invention are useful in the prevention and treatment of diseases caused by hepatitis C virus in animals, especially mammals, and most especially humans.

In providing the antibodies of the present invention to a recipient mammal, preferably a human, the dosage of administered antibodies will vary depending on such factors as the mammal's age, weight, height, sex, general medical condition, previous medical history, and the like.

In general, it will be advantageous to provide the recipient mammal with a dosage of antibodies in the range of from about 1 mg/kg body weight to about 10 mg/kg body weight of the mammal, although a lower or higher dose may be administered if found desirable. Such antibodies will normally be administered by intravenous or intramuscular route as an inoculum. The antibodies of the present invention are intended to be provided to the recipient subject in an amount sufficient to prevent, lessen or attenuate the severity, extent or duration of any existing infection.

The antibodies prepared by use of the nucleic acid sequences, viruses or polypeptides of the present invention are also highly useful for diagnostic purposes. For example, the antibodies can be used as in vitro

- 30 -

diagnostic agents to test for the presence of HCV in biological samples taken from animals, especially humans. Such assays include, but are not limited to, radioimmunoassays, EIA, fluorescence, Western blot analysis and ELISAs. In one such embodiment, the biological sample is contacted with antibodies of the present invention and a labeled second antibody is used to detect the presence of HCV to which the antibodies are bound.

Such assays may be, for example, a direct protocol (where the labeled first antibody is immunoreactive with the antigen, such as, for example, a polypeptide on the surface of the virus), an indirect protocol (where a labeled second antibody is reactive with the first antibody), a competitive protocol (such as would involve the addition of a labeled antigen), or a sandwich protocol (where both labeled and unlabeled antibody are used), as well as other protocols well known and described in the art.

In one embodiment, an immunoassay method would utilize an antibody specific for HCV envelope determinants and would further comprise the steps of contacting a biological sample with the HCV-specific antibody and then detecting the presence of HCV material in the test sample using one of the types of assay protocols as described above. Polypeptides and antibodies produced according to the present invention may also be supplied in the form of a kit, either present in vials as purified material, or present in compositions and suspended in suitable diluents as previously described.



- 31 -

°

In a preferred embodiment, such a diagnostic test kit for detection of HCV antigens in a test sample comprises in combination a series of containers, each container a reagent needed for such assay. Thus, one such  
5 container would contain a specific amount of HCV-specific antibody as already described, a second container would contain a diluent for suspension of the sample to be tested, a third container would contain a positive control  
10 and an additional container would contain a negative control. An additional container could contain a blank.

For all prophylactic, therapeutic and diagnostic uses, the antibodies of the invention and other reagents, plus appropriate devices and accessories, may be provided  
15 in the form of a kit so as to facilitate ready availability and ease of use.

The present invention also relates to the use of nucleic acid sequences and polypeptides of the present  
20 invention to screen potential antiviral agents for antiviral activity against HCV. Such screening methods are known by those of skill in the art. Generally, the antiviral agents are tested at a variety of  
25 concentrations, for their effect on preventing viral replication in cell culture systems which support viral replication, and then for an inhibition of infectivity or of viral pathogenicity (and a low level of toxicity) in an animal model system.

30 In one embodiment, animal cells (especially human cells) transfected with the nucleic acid sequences of the invention are cultured in vitro and the cells are treated with a candidate antiviral agent (a chemical,

35

- 32 -

peptide etc.) for antiviral activity by adding the candidate agent to the medium. The treated cells are then exposed, possibly under transfecting or fusing conditions known in the art, to the nucleic acid sequences of the present invention. A sufficient period of time would then be allowed to pass for infection to occur, following which the presence or absence of viral replication would be determined versus untreated control cells by methods known to those of ordinary skill in the art. Such methods include, but are not limited to, the detection of viral antigens in the cell, for example, by immunofluorescent procedures well known in the art; the detection of viral polypeptides by Western blotting using antibodies specific therefor; the detection of newly transcribed viral RNA within the cells by RT-PCR; and the detection of the presence of live, infectious virus particles by injection of cell culture medium or cell lysates into healthy, susceptible animals, with subsequent exhibition of the symptoms of HCV infection. A comparison of results obtained for control cells (treated only with nucleic acid sequence) with those obtained for treated cells (nucleic acid sequence and antiviral agent) would indicate, the degree, if any, of antiviral activity of the candidate antiviral agent. Of course, one of ordinary skill in the art would readily understand that such cells can be treated with the candidate antiviral agent either before or after exposure to the nucleic acid sequence of the present invention so as to determine what stage, or stages, of viral infection and replication said agent is effective against.

35

- 33 -

o

In an alternative embodiment, a protease such as NS3 protease produced from a nucleic acid sequence of the invention may be used to screen for protease inhibitors which may act as antiviral agents. The structural and nonstructural regions of the HCV genome, including nucleotide and amino acid locations, have been determined, for example, as depicted in Houghton, M. (1996), Fig. 1; and Major, M.E. et al. (1997), Table 1.

Such above-mentioned protease inhibitors may take the form of chemical compounds or peptides which mimic the known cleavage sites of the protease and may be screened using methods known to those of skill in the art (Houghton, M. (1996) and Major, M.E. et al. (1997)). For example, a substrate may be employed which mimics the protease's natural substrate, but which provides a detectable signal (e.g. by fluorimetric or colorimetric methods) when cleaved. This substrate is then incubated with the protease and the candidate protease inhibitor under conditions of suitable pH, temperature etc. to detect protease activity. The proteolytic activities of the protease in the presence or absence of the candidate inhibitor are then determined.

In yet another embodiment, a candidate antiviral agent (such as a protease inhibitor) may be directly assayed in vivo for antiviral activity by administering the candidate antiviral agent to a chimpanzee transfected with a nucleic acid sequence of the invention and then measuring viral replication in vivo via methods such as RT-PCR. Of course, the chimpanzee may be treated with the candidate agent either before or after transfection with

- 34 -

the infectious nucleic acid sequence so as to determine what stage, or stages, of viral infection and replication the agent is effective against.

The invention also provides that the nucleic acid sequences, viruses and polypeptides of the invention may be supplied in the form of a kit, alone or in the form of a pharmaceutical composition.

All scientific publication and/or patents cited herein are specifically incorporated by reference. The following examples illustrate various aspects of the invention but are in no way intended to limit the scope thereof.

#### EXAMPLES

##### **MATERIALS AND METHODS** **For Examples 1-4**

##### **Collection of Virus**

Hepatitis C virus was collected and used as a source for the RNA used in generating the cDNA clones according to the present invention. Plasma containing strain H77 of HCV was obtained from a patient in the acute phase of transfusion-associated non-A, non-B hepatitis (Feinstone et al (1981)). Strain H77 belongs to genotype 1a of HCV (Ogata et al (1991), Inchauspe et al (1991)). The consensus sequence for most of its genome has been determined (Kolyakov et al (1996), Ogata et al (1991), Inchauspe et al (1991) and Farci et al (1996)).

- 35 -

RNA Purification

Viral RNA was collected and purified by conventional means. In general, total RNA from 10  $\mu$ l of H77 plasma was extracted with the TRIzol system (GIBCO BRL). The RNA pellet was resuspended in 100  $\mu$ l of 10 mM dithiothreitol (DTT) with 5% (vol/vol) RNasin (20 - 40 units/ $\mu$ l) (available from Promega) and 10  $\mu$ l aliquots were stored at -80°C. In subsequent experiments RT-PCR was performed on RNA equivalent to 1  $\mu$ l of H77 plasma, which contained an estimated  $10^5$  genome equivalents (GE) of HCV (Yanagi et al (1996)).

Primers used in the RT-PCR process were deduced from the genomic sequences of strain H77 according to procedures already known in the art (see above) or else were determined specifically for use herein. The primers generated for this purpose are listed in Table 1.

- 36 -

Table 1. Oligonucleotides used for PCR amplification of strain H77 of HCV

	Designation	Sequence (5' → 3')*
5	H9261F	GGCTACAGCGGGGGGAGACATTTATCACAGC
	H3' X58R	TCATGCGGCTCACGGACCTTTCACAGCTAG
	H9282F	<b>GTCCAAGCTT</b> ATCACAGCGTGTCTCATGCCCCGGCCCCG
	H3' X45R	<b>CGTCTCTAGAGG</b> ACCTTTCACAGCTAGCCGTGACTAGGG
	H9375F	TGAAGGTTGGGGTAAACACTCCGGCCTCTTAGGCCATT
	H3' X-35R	ACATGATCTGCAGAGAGGCCAGTATCAGCACTCTC
	H9386F	<b>GTCCAAGCTTACGCGTAA</b> CACTCCGGCCTCCTTAAGCCATTCTCTG
10	H3' X-38R	<b>CGTCTCTAGACAT</b> GATCTGCAGAGAGGCCAGTATCAGCACTCTCTGC
	H1	<i>TTTTTTTTGCGGCCGCTAATACGACTCACTATAGCCAGCCCCCTGAT-</i>
		GGGGGCGACACTCCACCATG
	A1	ACTGTCTTCACGCAGAAAGCGTCTAGCCAT
	H9417R	<b>CGTCTCTAGACAGGAA</b> TGGCTTAAGAGGCCGGAGTGTTTACC
	* HCV sequences are shown in plain text, non-HCV-specific sequences are shown in boldface and artificial cleavage sites used for cDNA cloning are underlined. The core sequence of the T7 promoter in primer H1 is shown in italics.	
15	Primers for long RT-PCR were size-purified.	

### cDNA Synthesis

The RNA was denatured at 65°C for 2 min, and cDNA synthesis was performed in a 20 µl reaction volume with Superscript II reverse transcriptase (from GIBCO/BRL) at 42 °C for 1 hour using specific antisense primers as described previously (Tellier et al (1996)). The cDNA mixture was treated with RNase H and RNase T1 (GIBCO/BRL) for 20 min at 37 °C.

### Amplification and Cloning of the 3' UTR

The 3' UTR of strain H77 was amplified by PCR in two different assays. In both of these nested PCR reactions the first round of PCR was performed in a total volume of 50 µl in 1 x buffer, 250 µmol of each deoxynucleoside triphosphate (dNTP; Pharmacia), 20 pmol

- 37 -

°

5

10

15

20

25

30

35

each of external sense and antisense primers, 1  $\mu$ l of the Advantage KlenTaq polymerase mix (from Clontech) and 2  $\mu$ l of the final cDNA reaction mixture. In the second round of PCR, 5  $\mu$ l of the first round PCR mixture was added to 45  $\mu$ l of PCR mixture prepared as already described. Each round of PCR (35 cycles), which was performed in a Perkin Elmer DNA thermal cycler 480, consisted of denaturation at 94 °C for 1 min (in 1st cycle 1 min 30 sec), annealing at 60°C for 1 min and elongation at 68°C for 2 min. In one experiment a region from NS5B to the conserved region of the 3' UTR was amplified with the external primers H9261F and H3'X58R, and the internal primers H9282F and H3'X45R (Table 1). In another experiment, a segment of the variable region to the very end of the 3' UTR was amplified with the external primers H9375F and H3'X-35R, and the internal primers H9386F and H3'X-38R (Table 1, Fig. 1). Amplified products were purified with QIAquick PCR purification kit (from QIAGEN), digested with *Hind* III and *Xba* I (from Promega), purified by either gel electrophoresis or phenol/chloroform extraction, and then cloned into the multiple cloning site of plasmid pGEM-9zf(-) (Promega) or pUC19 (Pharmacia). Cloning of cDNA into the vector was performed with T4 DNA ligase (Promega) by standard procedures.

#### Amplification of Near Full-Length H77 Genomes by Long PCR

The reactions were performed in a total volume of 50  $\mu$ l in 1 x buffer, 250  $\mu$ mol of each dNTP, 10 pmol each of sense and antisense primers, 1  $\mu$ l of the Advantage

- 38 -

o

KlenTaq polymerase mix and 2  $\mu$ l of the cDNA reaction mixture (Tellier et al (1996)). A single PCR round of 35 cycles was performed in a Robocycler thermal cycler (from Stratagene), and consisted of denaturation at 99 °C for 35  
5 sec, annealing at 67 °C for 30 sec and elongation at 68 °C for 10 min during the first 5 cycles, 11 min during the next 10 cycles, 12 min during the following 10 cycles and 13 min during the last 10 cycles. To amplify the complete  
10 ORF of HCV by long RT-PCR we used the sense primers H1 or A1 deduced from the 5' UTR and the antisense primer H9417R deduced from the variable region of the 3' UTR (Table 1, Fig. 1).

15

#### Construction of Full-Length H77 cDNA Clones

The long PCR products amplified with H1 and H9417R primers were cloned directly into pGEM-9zf(-) after  
20 digestion with Not I and Xba I (from Promega) (as per Fig. 1). Two clones were obtained with inserts of the expected size, pH21<sub>r</sub> and pH50<sub>r</sub>. Next, the chosen 3' UTR was cloned into both pH21<sub>r</sub> and pH50<sub>r</sub> after digestion with  
25 Afl II and Xba I (New England Biolabs). DH5 $\alpha$  competent cells (GIBCO/BRL) were transformed and selected with LB agar plates containing 100  $\mu$ g/ml ampicillin (from SIGMA). Then the selected colonies were cultured in LB liquid  
30 containing ampicillin at 30°C for ~18-20 hrs (transformants containing full-length or near full-length cDNA of H77 produced a very low yield of plasmid when cultured at 37 °C or for more than 24 hrs). After small  
scale preparation (Wizard Plus Minipreps DNA Purification

35



- 39 -

Systems, Promega) each plasmid was retransformed to select a single clone, and large scale preparation of plasmid DNA was performed with a QIAGEN plasmid Maxi kit.

#### 5 Cloning of Long RT-PCR Products Into a Cassette Vector

To improve the efficiency of cloning, a vector with consensus 5' and 3' termini of HCV strain H77 was constructed (Fig. 1). This cassette vector (pCV) was  
10 obtained by cutting out the *Bam*HI fragment (nts 1358 - 7530 of the H77 genome) from pH50, followed by religation. Next, the long PCR products of H77 amplified with H1 and H9417R or A1 and H9417R primers were purified (Geneclean spin kit; BIO 101) and cloned into pCV after digestion  
15 with Age I and Afl II (New England Biolabs) or with *Pin* AI (isoschizomer of Age I) and *Bfr* I (isoschizomer of Afl II) (Boehringer Mannheim). Large scale preparations of the plasmids containing full-length cDNA of H77 were performed  
20 as described above.

#### Construction of H77 Consensus Chimeric cDNA Clone

A full-length cDNA clone of H77 with an ORF  
25 encoding the consensus amino acid sequence was constructed by making a chimera from four of the cDNA clones obtained above. This consensus chimera, pCV-H77C, was constructed in two ligation steps by using standard molecular  
30 procedures and convenient cleavage sites and involved first a two piece ligation and then a three piece ligation. Large scale preparation of pCV-H77C was performed as already described.

35

- 40 -

o

### In Vitro Transcription

Plasmids containing the full-length HCV cDNA were linearized with *Xba* I (from Promega), and purified by phenol/chloroform extraction and ethanol precipitation. A  
5 100  $\mu$ l reaction mixture containing 10  $\mu$ g of linearized plasmid DNA, 1 x transcription buffer, 1 mM ATP, CTP, GTP and UTP, 10mM DTT , 4% (v/v) RNasin (20-40 units/ $\mu$ l) and 2  
10  $\mu$ l of T7 RNA polymerase (Promega) was incubated at 37 °C for 2 hrs. Five  $\mu$ l of the reaction mixture was analyzed by agarose gel electrophoresis followed by ethidium bromide staining. The transcription reaction mixture was diluted with 400  $\mu$ l of ice-cold phosphate-buffered saline  
15 without calcium or magnesium, immediately frozen on dry ice and stored at -80 °C. The final nucleic acid mixture was injected into chimpanzees within 24 hrs.

### Intrahepatic Transfection of Chimpanzees

20 Laparotomy was performed and aliquots from two transcription reactions were injected into 6 sites of the exposed liver (Emerson et al (1992). Serum samples were collected weekly from chimpanzees and monitored for liver  
25 enzyme levels and anti-HCV antibodies. Weekly samples of 100  $\mu$ l of serum were tested for HCV RNA in a highly sensitive nested RT-PCR assay with AmpliTaq Gold (Perkin Elmer) (Yanagi et al (1996); Bukh et al (1992)). The  
30 genome titer of HCV was estimated by testing 10-fold serial dilutions of the extracted RNA in the RT-PCR assay (Yanagi et al (1996)). The two chimpanzees used in this

35

- 41 -

study were maintained under conditions that met all requirements for their use in an approved facility.

The consensus sequence of the complete ORF from HCV genomes recovered at week 2 post inoculation (p.i) was determined by direct sequencing of PCR products obtained in long RT-PCR with primers A1 and H9417R followed by nested PCR of 10 overlapping fragments. The consensus sequence of the variable region of the 3' UTR was determined by direct sequencing of an amplicon obtained in nested RT-PCR as described above. Finally, we amplified selected regions independently by nested RT-PCR with AmpliTaq Gold.

#### 15 Sequence Analysis

Both strands of DNA from PCR products, as well as plasmids, were sequenced with the ABI PRISM Dye Termination Cycle Sequencing Ready Reaction Kit using Taq DNA polymerase (Perkin Elmer) and about 100 specific sense and antisense sequence primers.

The consensus sequence of HCV strain H77 was determined in two different ways. In one approach, overlapping PCR products were directly sequenced, and amplified in nested RT-PCR from the H77 plasma sample. The sequence analyzed (nucleotides (nts) 35-9417) included the entire genome except the very 5' and 3' termini. In the second approach, the consensus sequence of nts 157-9384 was deduced from the sequences of 18 full-length cDNA clones.

- 42 -

## EXAMPLE 1

Variability in the sequence of the 3' UTR of HCV strain H77

5                   The heterogeneity of the 3' UTR was analyzed by  
cloning and sequencing of DNA amplicons obtained in nested  
RT-PCR. 19 clones containing sequences of the entire  
variable region, the poly U-UC region and the adjacent 19  
10 nt of the conserved region, and 65 clones containing  
sequences of the entire poly U-UC region and the first 63  
nts of the conserved region were analyzed. This analysis  
confirmed that the variable region consisted of 43 nts,  
including two conserved termination codons (Han et al  
15 (1992)). The sequence of the variable region was highly  
conserved within H77 since only 3 point mutations were  
found among the 19 clones analyzed. A poly U-UC region  
was present in all 84 clones analyzed. However, its  
20 length varied from 71-141 nts. The length of the poly U  
region was 9-103 nts, and that of the poly UC region was  
35-85 nts. The number of C residues increased towards the  
3' end of the poly UC region but the sequence of this  
region is not conserved. The first 63 nts of the  
25 conserved region were highly conserved among the clones  
analyzed, with a total of only 14 point mutations. To  
confirm the validity of the analysis, the 3' UTR was  
reamplified directly from a full-length cDNA clone of HCV  
30 (see below) by the nested-PCR procedure with the primers  
in the variable region and at the very 3' end of the HCV  
genome and cloned the PCR product. Eight clones had 1-7  
nt deletions in the poly U region. Furthermore, although

35

- 43 -

the C residues of the poly UC region were maintained, the spacing of these varied because of 1-2 nt deletions of U residues. These deletions must be artifacts introduced by PCR and such mistakes may have contributed to the heterogeneity originally observed in this region.

However, the conserved region of the 3' UTR was amplified correctly, suggesting that the deletions were due to difficulties in transcribing a highly repetitive sequence.

One of the 3' UTR clones was selected for engineering of full-length cDNA clones of H77. This clone had the consensus variable sequence except for a single point mutation introduced to create an Afl II cleavage site, a poly U-UC stretch of 81 nts with the most commonly observed UC pattern and the consensus sequence of the complete conserved region of 101 nts, including the distal 38 nts which originated from the antisense primer used in the amplification. After linearization with Xba I, the DNA template of this clone had the authentic 3' end.

#### EXAMPLE 2

##### The Entire Open Reading Frame of H77 Amplified in One Round of Long RT-PCR

It had been previously demonstrated that a 9.25 kb fragment of the HCV genome from the 5' UTR to the 3' end of NS5B could be amplified from  $10^4$  GE (genome equivalents) of H77 by a single round of long RT-PCR (Tellier et al (1996a)). In the current study, by optimizing primers and cycling conditions, the entire ORF of H77 was amplified in a single round of long RT-PCR with primers from the 5' UTR and the variable region of the 3'

- 44 -

UTR. In fact, 9.4 kb of the H77 genome (H product: from the very 5' end to the variable region of the 3' UTR) could be amplified from  $10^5$  GE or 9.3 kb (A product: from within the 5' UTR to the variable region of the 3' UTR) from  $10^4$  GE or  $10^5$  GE, in a single round of long RT-PCR (Fig. 2). The PCR products amplified from  $10^5$  GE of H77 were used for engineering full-length cDNA clones (see below).

## EXAMPLE 3

Construction of Multiple Full-Length  
cDNA Clones of H77 in a Single Step by  
Cloning of Long RT-PCR Amplicons Directly  
into a Cassette Vector with Fixed 5' and 3' Termini

Direct cloning of the long PCR products (H), which contained a 5' T7 promoter, the authentic 5' end, the entire ORF of H77 and a short region of the 3' UTR, into pGEM-9zf(-) vector by Not I and Xba I digestion was first attempted. However, among the 70 clones examined all but two had inserts that were shorter than predicted. Sequence analysis identified a second Not I site in the majority of clones, which resulted in deletion of the nts past position 9221. Only two clones (pH21<sub>r</sub> and pH50<sub>r</sub>) were missing the second Not I site and had the expected 5' and 3' sequences of the PCR product. Therefore, full-length cDNA clones (pH21 and pH50) were constructed by inserting the chosen 3' UTR into pH21<sub>r</sub> and pH50<sub>r</sub>, respectively. Sequence analysis revealed that clone pH21 had a complete full-length sequence of H77; this clone was tested for infectivity. The second clone, pH50, had one nt deletion

- 45 -

in the ORF at position 6365; this clone was used to make a cassette vector.

The complete ORF was amplified by constructing a cassette vector with fixed 5' and 3' termini as an intermediate of the full-length cDNA clones. This vector (pCV) was constructed by digestion of clone pH50 with *Bam*HI, followed by religation, to give a shortened plasmid readily distinguished from plasmids containing the full-length insert. Attempts to clone long RT-PCR products (H) into pCV by *Age* I and *Afl* II yielded only 1 of 23 clones with an insert of the expected size. In order to increase the efficiency of cloning, we repeated the procedure but used *Pin* A I and *Bfr* I instead of the respective isoschizomers *Age* I and *Afl* II. By this protocol, 24 of 31 H clones and 30 of 35 A clones had the full-length cDNA of H77 as evaluated by restriction enzyme digestion. A total of 16 clones, selected at random, were each retransformed, and individual plasmids were purified and completely sequenced.

#### EXAMPLE 4

##### Demonstration of Infectious Nature of Transcripts of a cDNA Clone Representing the Consensus Sequence of Strain H77

A consensus chimera was constructed from 4 of the full-length cDNA clones with just 2 ligation steps. The final construct, pCV-H77C, had 11 nt differences from the consensus sequence of H77 in the ORF (Fig. 3). However, 10 of these nucleotide differences represented silent mutations. The chimeric clone differed from the

35

- 46 -

o

consensus sequence at only one amino acid [L instead of F at position 790]. Among the 18 ORFs analyzed above, the F residue was found in 11 clones and the L residue in 7 clones. However, the L residue was dominant in other isolates of genotype 1a, including a first passage of H77 in a chimpanzee (Inchauspe et al (1991)).

To test the infectivity of the consensus chimeric clone of H77 intrahepatic transfection of a chimpanzee was performed. The pCV-H77C clone was linearized with Xba I and transcribed *in vitro* by T7 RNA polymerase (Fig. 2). The transcription mixture was next injected into 6 sites of the liver of chimpanzee 1530. The chimpanzee became infected with HCV as measured by detection of  $10^2$  GE/ml of viral genome at week 1 p.i. Furthermore, the HCV titer increased to  $10^4$  GE/ml at week 2 p.i., and reached  $10^6$  GE/ml by week 8 p.i. The viremic pattern observed in the early phase of the infection with the recombinant virus was similar to that observed in chimpanzees inoculated intravenously with strain H77 or other strains of HCV (Shimizu (1990)).

The sequence of the HCV genomes from the serum sample collected at week 2 p.i. was analyzed. The consensus sequence of nts 298-9375 of the recovered genomes was determined by direct sequencing of PCR products obtained in long RT-PCR followed by nested PCR of 10 overlapping fragments. The identity to clone pCV-H77C sequence was 100%. The consensus sequence of nts 96-291, 1328-1848, 3585-4106, 4763-5113 and 9322-9445 was determined from PCR products obtained in different nested RT-PCR assays. The identity of these sequences with pCV-

35



- 47 -

0  
H77C was also 100%. These latter regions contained 4  
mutations unique to the consensus chimera, including the  
artificial Afl II cleavage site in the 3' UTR. Therefore,  
RNA transcripts of this clone of HCV were infectious.

5           The infectious nature of the consensus chimera  
indicates that the regions of the 5' and 3' UTRs  
incorporated into the cassette vector do not destroy  
viability. This makes it highly advantageous to use the  
10 cassette vector to construct infectious cDNA clones of  
other HCV strains when the consensus sequence for each ORF  
is inserted.

          In addition, two complete full-length clones  
15 (dubbed pH21 and pCV-H11) constructed were not infectious,  
as shown by intrahepatic injection of chimpanzees with the  
corresponding RNA transcripts. Thus, injection of the  
transcription mixture into 3 sites of the exposed liver  
resulted in no observable HCV replication and weekly serum  
20 samples were negative for HCV RNA at weeks 1 - 17 p.i. in  
a highly sensitive nested RT-PCR assay. The cDNA template  
injected along with the RNA transcripts was also not  
detected in this assay.

25           Moreover, the chimpanzee remained negative for  
antibodies to HCV throughout the follow-up. Subsequent  
sequence analysis revealed that 7 of 16 additional clones  
were defective for polyprotein synthesis and all clones  
had multiple amino acid mutations compared with the  
30 consensus sequence of the parent strain. For example,  
clone pH21, which was not infectious, had 7 amino acid  
substitutions in the entire predicted polyprotein compared  
with the consensus sequence of H77 (Fig. 3). The most  
35

- 48 -

notable mutation was at position 1026, which changed L to Q, altering the cleavage site between NS2 and NS3 (Reed (1995)). Clone pCV-H11, also non-infectious, had 21 amino acid substitutions in the predicted polyprotein compared with the consensus sequence of H77 (Fig. 3). The amino acid mutation at position 564 eliminated a highly conserved C residue in the E2 protein (Okamoto (1992a)).

#### EXAMPLE 4A

The chimpanzee of Example 4, designated 1530, was monitored out to 32 weeks p.i. for serum enzyme levels (ALT) and the presence of anti-HCV antibodies, HCV RNA, and liver histopathology. The results are shown in Figure 18B.

A second chimp, designated 1494, was also transfected with RNA transcripts of the pCV-H77C clone and monitored out to 17 weeks p.i. for the presence of anti-HCV antibodies, HCV RNA and elevated serum enzyme levels. The results are shown in Figure 18A.

#### MATERIALS AND METHODS for Examples 5-10

##### Source Of HCV Genotype 1b

An infectious plasma pool (second chimpanzee passage) containing strain HC-J4, genotype 1b, was prepared from acute phase plasma of a chimpanzee experimentally infected with serum containing HC-J4/91 (Okamoto et al., 1992b). The HC-J4/91 sample was obtained from a first chimpanzee passage during the chronic phase of hepatitis C about 8 years after experimental infection.

- 49 -

The consensus sequence of the entire genome, except for the very 3' end, was determined previously for HC-J4/91 (Okamoto et al., 1992b).

#### 5     Preparation Of HCV RNA

Viral RNA was extracted from 100  $\mu$ l aliquots of the HC-J4 plasma pool with the TRIzol system (GIBCO BRL). The RNA pellets were each resuspended in 10  $\mu$ l of 10 mM  
10     dithiothreitol (DTT) with 5% (vol/vol) RNasin (20-40 units/ $\mu$ l) (Promega) and stored at -80°C or immediately used for cDNA synthesis.

#### 15     Amplification And Cloning Of The 3' UTR

A region spanning from NS5B to the conserved region of the 3' UTR was amplified in nested RT-PCR using the procedure of Yanagi et al., (1997).

20     In brief, the RNA was denatured at 65°C for 2 minutes, and cDNA was synthesized at 42°C for 1 hour with Superscript II reverse transcriptase (GIBCO BRL) and primer H3'X58R (Table 1) in a 20  $\mu$ l reaction volume. The  
25     cDNA mixture was treated with RNase H and RNase T1 (GIBCO BRL) at 37°C for 20 minutes. The first round of PCR was performed on 2  $\mu$ l of the final cDNA mixture in a total volume of 50  $\mu$ l with the Advantage cDNA polymerase mix  
30     (Clontech) and external primers H9261F (Table 1) and H3'X58R (Table 1). In the second round of PCR [internal primers H9282F (Table 1) and H3'X45R (Table 1)], 5  $\mu$ l of the first round PCR mixture was added to 45  $\mu$ l of the PCR reaction mixture. Each round of PCR (35 cycles), was  
35

- 50 -

performed in a DNA thermal cycler 480 (Perkin Elmer) and consisted of denaturation at 94°C for 1 minute (1st cycle: 1 minute 30 sec), annealing at 60°C for 1 minute and elongation at 68°C for 2 minutes. After purification with QIAquick PCR purification kit (QIAGEN), digestion with *HindIII* and *XbaI* (Promega), and phenol/chloroform extraction, the amplified products were cloned into pGEM-9zf(-) (Promega) (Yanagi et al., 1997).

#### Amplification And Cloning Of The Entire ORF

A region from within the 5' UTR to the variable region of the 3' UTR of strain HC-J4 was amplified by long RT-PCR (Fig. 1) (Yanagi et al., 1997). The cDNA was synthesized at 42°C for 1 hour in a 20 µl reaction volume with Superscript II reverse transcriptase and primer J4-9405R (5'-GCCTATTGGCCTGGAGTGGTTAGCTC-3'), and treated with RNases as above. The cDNA mixture (2 µl) was amplified by long PCR with the Advantage cDNA polymerase mix and primers A1 (Table 1) (Bukh et al., 1992; Yanagi et al., 1997) and J4-9398R (5'-

**AGGATGGCCTTAAGGCCTGGAGTGGTTAGCTCCCCGTTCA**-3'). Primer J4-9398R contained extra bases (**bold**) and an artificial *AflII* cleavage site (underlined). A single PCR round was performed in a Robocycler thermal cycler (Stratagene), and consisted of denaturation at 99°C for 35 seconds, annealing at 67°C for 30 seconds and elongation at 68°C for 10 minutes during the first 5 cycles, 11 minutes during the next 10 cycles, 12 minutes during the following 10 cycles and 13 minutes during the last 10 cycles.

- 51 -

After digesting the long PCR products obtained from strain HC-J4 with *PinAI* (isoschizomer of *AgeI*) and *BfrI* (isoschizomer of *AflIII*) (Boehringer Mannheim), attempts were made to clone them directly into a cassette vector (pCV), which contained the 5' and 3' termini of strain H77 (Figure 1) but no full-length clones were obtained. Accordingly, to improve the efficiency of cloning, the PCR product was further digested with *BglIII* (Boehringer Mannheim) and the two resultant genome fragments [L fragment: *PinAI/BglIII*, nts 156 - 8935; S fragment: *BglIII/BrfI*, nts 8936 - 9398] were separately cloned into pCV (Figure 6).

DH5 $\alpha$  competent cells (GIBCO BRL) were transformed and selected on LB agar plates containing 100  $\mu$ g/ml ampicillin (SIGMA) and amplified in LB liquid cultures at 30°C for 18-20 hours.

Sequence analysis of 9 plasmids containing the S fragment (miniprep samples) and 9 plasmids containing the L fragment (maxiprep samples) were performed as described previously (Yanagi et al., 1997). Three L fragments, each encoding a distinct polypeptide, were cloned into pCV-J4S9 (which contained an S fragment encoding the consensus amino acid sequence of HC-J4) to construct three chimeric full-length HCV cDNAs (pCV-J4L2S, pCV-J4L4S and pCV-J4L6S) (Fig. 6). Large scale preparation of each clone was performed as described previously with a QIAGEN plasmid Maxi kit (Yanagi et al., 1997) and the authenticity of each clone was confirmed by sequence analysis.

- 52 -

### Sequence Analysis

Both strands of DNA were sequenced with the ABI PRISM Dye Termination Cycle Sequencing Ready Reaction Kit using Taq DNA polymerase (Perkin Elmer) and about 90 specific sense and antisense primers. Analyses of genomic sequences, including multiple sequence alignments and tree analyses, were performed with GeneWorks (Oxford Molecular Group) (Bukh et al., 1995).

The consensus sequence of strain HC-J4 was determined by direct sequencing of PCR products (nts 11 - 9412) and by sequence analysis of multiple cloned L and S fragments (nts 156 - 9371). The consensus sequence of the 3' UTR (3' variable region, polypyrimidine tract and the first 16 nucleotides of the conserved region) was determined by analysis of 24 cDNA clones.

### Intrahepatic Transfection Of A Chimpanzee With Transcribed RNA

Two in vitro transcription reactions were performed with each of the three full-length clones. In each reaction 10  $\mu$ g of plasmid DNA linearized with Xba I (Promega) was transcribed in a 100  $\mu$ l reaction volume with T7 RNA polymerase (Promega) at 37°C for 2 hours as described previously (Yanagi et al., 1997). Five  $\mu$ l of the final reaction mixture was analyzed by agarose gel electrophoresis and ethidium bromide staining (Fig. 5). Each transcription mixture was diluted with 400  $\mu$ l of ice-cold phosphate-buffered saline without calcium or magnesium and then the two aliquots from the same cDNA

35

- 53 -

clone were combined, immediately frozen on dry ice and stored at -80°C. Within 24 hours after freezing the transcription mixtures were injected into the chimpanzee by percutaneous intrahepatic injection that was guided by ultrasound. Each inoculum was individually injected (5-6 sites) into a separate area of the liver to prevent complementation or recombination. The chimpanzee was maintained under conditions that met all requirements for its use in an approved facility.

Serum samples were collected weekly from the chimpanzee and monitored for liver enzyme levels and anti-HCV antibodies. Weekly samples of 100 µl of serum were tested for HCV RNA in a sensitive nested RT-PCR assay (Bukh et al., 1992, Yanagi et al., 1996) with AmpliTaq Gold DNA polymerase. The genome equivalent (GE) titer of HCV was determined by testing 10-fold serial dilutions of the extracted RNA in the RT-PCR assay (Yanagi et al., 1996) with 1 GE defined as the number of HCV genomes present in the highest dilution which was positive in the RT-nested PCR assay.

To identify which of the three clones was infectious in vivo, the NS3 region (nts 3659 - 4110) from the chimpanzee serum was amplified in a highly sensitive and specific nested RT-PCR assay with AmpliTaq Gold DNA polymerase and the PCR products were cloned with a TA cloning kit (Invitrogen). In addition, the consensus sequence of the nearly complete genome (nts 11 - 9441) was determined by direct sequencing of overlapping PCR products.

- 54 -

EXAMPLE 5Sequence Analysis Of Infectious Plasma Pool  
Of Strain HC-J4 Used As The Cloning Source

5

As an infectious cDNA clone of a genotype 1a strain of HCV had been obtained only after the ORF was engineered to encode the consensus polypeptide (Kolykhalov et al., 1997; Yanagi et al., 1997), a detailed sequence analysis of the cloning source was performed to determine the consensus sequence prior to constructing an infectious cDNA clone of a 1b genotype.

15

A plasma pool of strain HC-J4 was prepared from acute phase plasmapheresis units collected from a chimpanzee experimentally infected with HC-J4/91 (Okamoto et al., 1992b). This HCV pool had a PCR titer of  $10^4$  -  $10^5$  GE/ml and an infectivity titer of approximately  $10^3$  chimpanzee infectious doses per ml.

20

The heterogeneity of the 3' UTR of strain HC-J4 was determined by analyzing 24 clones of nested RT-PCR product. The consensus sequence was identical to that previously published for HC-J4/91 (Okamoto et al., 1992b), except at position 9407 (see below). The variable region consisted of 41 nucleotides (nts. 9372 - 9412), including two in-frame termination codons. Furthermore, its sequence was highly conserved except at positions 9399 (19 A and 5 T clones) and 9407 (17 T and 7 A clones). The poly U-UC region varied slightly in composition and greatly in length (31-162 nucleotides). In the conserved region, the first 16 nucleotides of 22 clones were identical to those previously published for other genotype

35



- 55 -

1 strains, whereas two clones each had a single point mutation. These data suggested that the structural organization at the 3' end of HC-J4 was similar to that of the infectious clone of a genotype 1a strain of Yanagi et al (1997).

Next, the entire ORF of HC-J4 was amplified in a single round of long RT-PCR (Figure 5). The original plan was to clone the resulting PCR products into the *Pvu*AI and *Bst*FI site of a HCV cassette vector (pCV), which had fixed 5' and 3' termini of genotype 1a (Yanagi et al., 1997) but since full-length clones were not obtained, two genome fragments (L and S) derived from the long RT-PCR products (Figure 6) were separately subcloned into pCV.

To determine the consensus sequence of the ORF, the sequence of 9 clones each of the L fragment (pCV-J4L) and of the S fragment (pCV-J4S) was determined and quasispecies were found at 275 nucleotide (3.05 %) and 78 amino acid (2.59 %) positions, scattered throughout the 9030 nts (3010 aa) of the ORF (Figure 7). Of the 161 nucleotide substitutions unique to a single clone, 71% were at the third position of the codon and 72 % were silent.

Each of the nine L clones represented the near complete ORF of an individual genome. The differences among the L clones were 0.30 - 1.53% at the nucleotide and 0.31 - 1.47% at the amino acid level (Figure 8). Two clones, L1 and L7, had a defective ORF due to a single nucleotide deletion and a single nucleotide insertion, respectively. Even though the HC-J4 plasma pool was obtained in the early acute phase, it appeared to contain

35

- 56 -

at least three viral species (Figure 9). Species A contained the L1, L2, L6, L8 and L9 clones, species B the L3, L7 and L10 clones and species C the L4 clone. Although each species A clone was unique all A clones differed from all B clones at the same 20 amino acid sites and at these positions, species C had the species A sequence at 14 positions and the species B sequence at 6 positions (Figure 7).

Okamoto and coworkers (Okamoto et al., 1992b) previously determined the nearly complete genome consensus sequence of strain HC-J4 in acute phase serum of the first chimpanzee passage (HC-J4/83) as well as in chronic phase serum collected 8.2 years later (HC-J4/91). In addition, they determined the sequence of amino acids 379 to 413 (including HVR1) and amino acids 468 to 486 (including HVR2) of multiple individual clones (Okamoto et al., 1992b).

It was found by the present inventors that the sequences of individual genomes in the plasma pool collected from a chimpanzee inoculated with HC-J4/91 were all more closely related to HC-J4/91 than to HC-J4/83 (Figures 8, 9) and contained HVR amino acid sequences closely related to three of the four viral species previously found in HC-J4/91 (Figure 10).

Thus, the data presented herein demonstrate the occurrence of the simultaneous transmission of multiple species to a single chimpanzee and clearly illustrates the difficulties in accurately determining the evolution of HCV over time since multiple species with significant changes throughout the HCV genome can be present from the

- 57 -

onset of the infection. Accordingly, infection of chimpanzees with monoclonal viruses derived from the infectious clones described herein will make it possible to perform more detailed studies of the evolution of HCV in vivo and its importance for viral persistence and pathogenesis.

#### EXAMPLE 6

##### Determination Of The Consensus Sequence Of HC-J4 In The Plasma Pool

The consensus sequence of nucleotides 156-9371 of HC-J4 was determined by two approaches. In one approach, the consensus sequence was deduced from 9 clones of the long RT-PCR product. In the other approach the long RT-PCR product was reamplified by PCR as overlapping fragments which were sequenced directly. The two "consensus" sequences differed at 31 (0.34%) of 9216 nucleotide positions and at 11 (0.37%) of 3010 deduced amino acid positions (Figure 7). At all of these positions a major quasispecies of strain HC-J4 was found in the plasma pool. At 9 additional amino acid positions the cloned sequences displayed heterogeneity and the direct sequence was ambiguous (Figure 7). Finally, it should be noted that there were multiple amino acid positions at which the consensus sequence obtained by direct sequencing was identical to that obtained by cloning and sequencing even though a major quasispecies was detected (Figure 7).

For positions at which the two "consensus" sequences of HC-J4 differed, both amino acids were

- 58 -

included in a composite consensus sequence (Figure 7). However, even with this allowance, none of the 9 L clones analyzed (aa 1 - 2864) had the composite consensus sequence: two clones did not encode the complete polypeptide and the remaining 7 clones differed from the consensus sequence by 3 - 13 amino acids (Figure 7).

#### EXAMPLE 7

##### Construction Of Chimeric Full-Length cDNA Clones Containing The Entire ORF Of HC-J4

The cassette vector used to clone strain H77 was used to construct an infectious cDNA clone containing the ORF of a second subtype.

In brief, three full-length cDNA clones were constructed by cloning different L fragments into the *PinAI/BglIII* site of pCV-J4S9, the cassette vector for genotype 1a (Figure 6), which also contained an S fragment encoding the consensus amino acid sequence of HC-J4. Therefore, although the ORF was from strain HC-J4, most of the 5' and 3' terminal sequences originated from strain H77. As a result, the 5' and 3' UTR were chimeras of genotypes 1a and 1b (Figure 11).

The first 155 nucleotides of the 5' UTR were from strain H77 (genotype 1a), and differed from the authentic sequence of HC-J4 (genotype 1b) at nucleotides 11, 12, 13, 34 and 35. In two clones (pCV-J4L2S, pCV-J4L6S) the rest of the 5' UTR had the consensus sequence of HC-J4, whereas the third clone (pCV-J4L4S) had a single nucleotide insertion at position 207. In all 3 clones the first 27 nucleotides of the 3' variable region of the 3'

- 59 -

o

UTR were identical with the consensus sequence of HC-J4. The remaining 15 nucleotides of the variable region, the poly U-UC region and the 3' conserved region of the 3' UTR had the same sequence as an infectious clone of strain H77 (Figure 11).

None of the three full-length clones of HC-J4 had the ORF composite consensus sequence (Figures 7, 12). The pCV-J4L6S clone had only three amino acid changes: Q for R at position 231 (E1), V for A at position 937 (NS2) and T for S at position 1215 (NS3). The pCV-J4L4S clone had 7 amino acid changes, including a change at position 450 (E2) that eliminated a highly conserved N-linked glycosylation site (Okamoto et al., 1992a). Finally, the pCV-J4L2S clone had 9 amino acid changes compared with the consensus sequence of HC-J4. A change at position 304 (E1) mutated a highly conserved cysteine residue (Bukh et al., 1993; Okamoto et al., 1992a).

20

#### EXAMPLE 8

##### Transfection Of A Chimpanzee By In Vitro Transcripts Of A Chimeric cDNA

The infectivity of the three chimeric HCV clones was determined by ultra-sound-guided percutaneous intrahepatic injection into the liver of a chimpanzee of the same amount of cDNA and transcription mixture for each of the clones (Figure 5). This procedure is a less invasive procedure than the laparotomy procedure utilized by Kolykhalov et al. (1997) and Yanagi et al. (1997) and should facilitate in vivo studies of cDNA clones of HCV in

35

- 60 -

chimpanzees since percutaneous procedures, unlike laparotomy, can be performed repeatedly.

As shown in Figure 13, the chimpanzee became infected with HCV as measured by increasing titers of  $10^2$  GE/ml at week 1 p.i.,  $10^3$  GE/ml at week 2 p.i. and  $10^4$  -  $10^5$  GE/ml during weeks 3 to 10 p.i.

The viremic pattern found in the early phase of the infection was similar to that observed for the recombinant H77 virus in chimpanzees (Bukh et al., unpublished data; Kolykhalov et al., 1997; Yanagi et al., 1997). The chimpanzee transfected in the present study was chronically infected with hepatitis G virus (HGV/GBV-C) (Bukh et al., 1998) and had a titer of  $10^6$  GE/ml at the time of HCV transfection. Although HGV/GBV-C was originally believed to be a hepatitis virus, it does not cause hepatitis in chimpanzees (Bukh et al., 1998) and may not replicate in the liver (Laskus et al., 1997). The present study demonstrated that an ongoing infection of HGV/GBV-C did not prevent acute HCV infection in the chimpanzee model.

However, to identify which of the three full-length HC-J4 clones were infectious, the NS3 region (nts. 3659 - 4110) of HCV genomes amplified by RT-PCR from serum samples taken from the infected chimpanzee during weeks 2 and 4 post-infection (p.i.) were cloned and sequenced. As the PCR primers were a complete match with each of the original three clones, this assay should not have preferentially amplified one virus over another. Sequence analysis of 26 and 24 clones obtained at weeks 2 and 4

- 61 -

p.i., respectively, demonstrated that all originated from the transcripts of pCV-J4L6S.

Moreover, the consensus sequence of PCR products of the nearly complete genome (nts. 11-9441), amplified from serum obtained during week 2 p.i., was identical to the sequence of pCV-J4L6S and there was no evidence of quasispecies. Thus, RNA transcripts of pCV-J4L6S, but not of pCV-J4L2S or pCV-J4L4S, were infectious in vivo. The data in Figure 13 is therefore the product of the transfection of RNA transcripts of pCV-J4L6S.

In addition, the chimeric sequences of genotypes 1a and 1b in the UTRs were maintained in the infected chimpanzee. The consensus sequence of nucleotides 11 - 341 of the 5' UTR and the variable region of the 3' UTR, amplified from serum obtained during weeks 2 and 4 p.i., had the expected chimeric sequence of genotypes 1a and 1b (Fig. 11). Also three of four clones of the 3' UTR obtained at week 2 p.i. had the chimeric sequence of the variable region, whereas a single substitution was noted in the fourth clone. However, in all four clones the poly U region was longer (2-12 nts) than expected. Also, extra C and G residues were observed in this region. For the most part, the number of C residues in the poly UC region was maintained in all clones although the spacing varied. As shown previously, variations in the number of U residues can reflect artifacts introduced during PCR amplification (Yanagi et al., 1997). The sequence of the first 19 nucleotides of the conserved region was maintained in all four clones. Thus, with the exception of the poly U-UC region, the genomic sequences recovered

35

- 62 -

from the infected chimpanzee were exactly those of the chimeric infectious clone pCV-J4BL6S.

The results presented in Figure 13 therefore demonstrate that HCV polypeptide sequences other than the consensus sequence can be infectious and that a chimeric genome containing portions of the H77 termini could produce an infectious virus. In addition, these results showed for the first time that it is possible to make infectious viruses containing 5' and 3' terminal sequences specific for two different subtypes of the same major genotype of HCV.

#### EXAMPLE 9

##### Construction Of A Chimeric 1a/1b Infectious Clone

A chimeric 1a/1b infectious clone in which the structural region of the genotype 1b infectious clone is inserted into the 1a clone of Yanagi et al. (1997) is constructed by following the protocol shown in Figure 15. The resultant chimera contains nucleotides 156-2763 of the 1b clone described herein inserted into the 1a clone of Figures 4A-4F. The sequences of the primers shown in Figure 15 which are used in constructing this chimeric clone, designated pH77CV-J4, are presented below.



- 63 -

1. H2751S (Cla I/Nde I)  
CGT CAT CGA TCC TCA GCG GGC ATA TGC ACT GGA CAC GGA
2. H2870R  
CAT GCA CCA GCT GAT ATA GCG CTT GTA ATA TG
- 5 3. H7851S  
TCC GTA GAG GAA GCT TGC AGC CTG ACG CCC
4. H9173 R(P-M)  
GTA CTT GCC ACA TAT AGC AGC CCT GCC TCC TCT G
- 10 5. H9140S (P-M)  
CAG AGG AGG CAG GGC TGC TAT ATG TGG CAA GTA C
6. H9417R  
CGT CTC TAG ACA GGA AAT GGC TTA AGA GGC CGG AGT GTT  
TAC C
- 15 7. J4-2271S  
TGC AAT TGG ACT CGA GGA GAG CGC TGT AAC TTG GAG
8. J4-2776R (Nde I)  
CGG TCC AAG GCA TAT GCT CGT GGT GGT AAC GCC AG

Transcripts of the chimeric 1a/1b clone (whose  
 20 sequence is shown in Figures 16A-16F) are then produced  
 and transfected into chimpanzees by the methods described  
 in the Materials and Methods section herein and the  
 transfected animals are then be subjected to biochemical  
 25 (ALT levels), histopathological and PCR analyses to  
 determine the infectivity of the chimeric clone.

#### EXAMPLE 10

##### Construction of 3' Deletion Mutants 30 Of The 1a Infectious Clone pCV-H77C

Seven constructs having various deletions in the  
 3' untranslated region (UTR) of the 1a infectious clone  
 pCV-H77C were constructed as described in Figures 17A-17B.

35

- 64 -

The 3' untranslated sequence remaining in each of the seven constructs following their respective deletions is shown in Figures 17A-17B.

Construct pCV-H77C(-98X) containing a deletion of the 3'-most 98 nucleotide sequences in the 3'-UTR was transcribed in vitro according to the methods described herein and 1 ml of the diluted transcription mixture was percutaneously transfected into the liver of a chimpanzee with the aid of ultrasound. After three weeks, the transfection was repeated. The chimpanzee was observed to be negative for hepatitis C virus replication as measured by RT-PCR assay for 5 weeks after transfection. These results demonstrate that the deleted 98 nucleotide 3'-UTR sequence was critical for production of infectious HCV and appear to contradict the reports of Dash et al. (1996) and Yoo et al. (1995) who reported that RNA transcripts from cDNA clones of HCV-1 and HCV-N lacking the terminal 98 conserved nucleotides at the very 3' end of the 3'-UTR resulted in viral replication after transfection into human hematoma cell lines.

Transcripts of the (-42X) mutant (Figure 17C) were also produced and transfected into a chimpanzee and transcripts of the other five deletion mutants shown in Figures 17D-17G) are to be produced and transfected into chimpanzees by the methods described herein. All transfected animals are to then be assayed for viral replication via RT-PCR.

- 65 -

Discussion

In two recent reports on transfection of chimpanzees, only those clones engineered to have the independently determined and slightly different consensus amino acid sequence of the polypeptide of strain H77 were infectious (Kolykhalov et al., 1997; Yanagi et al., 1997). Although the two infectious clones differed at four amino acid positions, these differences were represented in a major component of the quasispecies of the cloning source. In the present study, a single consensus sequence of strain HC-J4 could not be defined because the consensus sequence obtained by two different approaches (direct sequencing and sequencing of cloned products) differed at 20 amino acid positions, even though the same genomic PCR product was analyzed. The infectious clone differed at two positions from the composite amino acid consensus sequence, from the sequence of the 8 additional HC-J4 clones analyzed in this study and from published sequences of earlier passage samples. An additional amino acid differed from the composite consensus sequence but was found in two other HC-J4 clones analyzed in this study. The two non-infectious full-length clones of HC-J4 differed from the composite consensus sequence by only 7 and 9 amino acid differences. However, since these clones had the same termini as the infectious clone (except for a single nucleotide insertion in the 5' UTR of pCV-J4L4S), one or more of these amino acid changes in each clone was apparently deleterious for the virus.

- 66 -

o

It was also found in the present study that HC-J4, like other strains of genotype 1b (Kolykhalov et al., 1996; Tanaka et al., 1996; Yamada et al., 1996), had a poly U-UC region followed by a terminal conserved element. The poly U-UC region appears to vary considerably so it was not clear whether changes in this region would have a significant effect on virus replication. On the other hand, the 3' 98 nucleotides of the HCV genome were previously shown to be identical among other strains of genotypes 1a and 1b (Kolykhalov et al., 1996; Tanaka et al., 1996). Thus, use of the cassette vector would not alter this region except for addition of 3 nucleotides found in strain H77 between the poly UC region and the 3' 98 conserved nucleotides.

In conclusion, an infectious clone representing a genotype 1b strain of HCV has been constructed. Thus, it has been demonstrated that it was possible to obtain an infectious clone of a second strain of HCV. In addition, it has been shown that a consensus amino acid sequence was not absolutely required for infectivity and that chimeras between the UTRs of two different genotypes could be viable.

30

35

- 67 -

REFERENCES

- Alter, M. J. (1997). Hepatology 26, 62S-65S.
- Blight, K. J. and Rice, C. M. (1997). J. Virol. 71,  
5 7345-7352.
- Brechot, C. (1997). Hepatology 25, 772-774.
- Bukh, J., et al. (1992). Proc. Natl. Acad. Sci. USA  
89, 187-191.
- Bukh, J., et al. (1993). Proc. Natl. Acad. Sci. USA  
10 90, 8234-8238.
- Bukh, J., et al. (1995). Semin. Liver Dis. 15,  
41-63.
- Bukh, J., et al. (1997). Genetic heterogeneity of  
15 hepatitis C virus and related viruses. In "Viral  
Hepatitis and Liver Disease, Proceedings of IX Triennial  
International Symposium on Viral Hepatitis and Liver  
Disease, Rome, Italy, 1996" (M. Rizzetto, R. H. Purcell,  
J. L. Gerin and G. Verme, Eds.), pp. 167-175. Edizioni  
20 Minerva Medica, Turin.
- Bukh, J., et al. (1998). Experimental infection of  
chimpanzees with hepatitis G virus and genetic analysis of  
the virus. J. Infect. Dis. (in press).
- 25 Choo, Q.-L., et al. (1991). Diversity of the  
hepatitis C virus. Proc. Natl. Acad. Sci. USA 88,  
2451-2455.
- Dash, S., et al. (1997). Am. J. Pathol. 151, 363-  
30 373.
- Emerson, S.U. et al (1992) J. Virol., 66:6649-6654.
- Farci, P. and Purcell, R. H. (1993). Natural history  
and experimental models. In "Viral Hepatitis: Scientific  
Basis and Clinical Management". (A. J. Zuckerman and H.
- 35

- 68 -

C. Thomas, Eds.). pp. 241-267. Churchill Livingstone, Edinburgh.

Farci, P., et al. (1994). Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated in vitro neutralization. Proc. Natl. Acad. Sci. USA 91, 7792-7796.

Farci, P., et al. (1996). Proc. Natl. Acad. Sci. USA 93, 15394-15399.

Farci, P., et al. (1997). Springer Semin. Immunopathol. 19, 5-26.

Fausto, N. (1997). Am. J. Pathol. 151, 361.

Feinstone, S.M. et al (1981) J. Infect. Dis. 144:588-598.

Fried, M. W. and Hoofnagle, J. H. (1995). Semin. Liver Dis. 15, 82-91.

Han, J.M. et al (1992) Nuc. Acids Res., 20:3250.

Hijikata, M., et al. (1991). Biochem. Biophys. Res. Commun. 175, 220-228.

Honda, M., et al. (1996). RNA 2, 955-968.

Hoofnagle, J. H. (1997). Hepatology 26, 15S-20S.

Houghton, M. (1996). Hepatitis C viruses. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven Publishers, Philadelphia.

Inchauspe, G. et al. (1991) Proc. Natl. Acad. Sci. U.S.A., 88:10292-10296.

Ito, T. and Lai, M. M. C. (1997). J. Virol. 71, 8698-8706.

Kato, N., et al. (1990). Proc. Natl. Acad. Sci. USA 87, 9524-9528.

35

- 69 -

o

Kolykhalov, A. A., Feinstone, S. M. and Rice, C. M.  
(1996). J. Virol. 70, 3363-3371.

Kolykhalov, A. A., et al. (1997). Science 277,  
570-574.

5

Krieg, A.M. et al. (1995) Nature, 374:546.

Krieg, A.M. et al. (1996) J. Lab. Clin. Med.,  
128:128.

Laskus, T., et al. (1997). J. Virol. 71, 7804-7806.

10

Major, M.E. and Feinstone, S.M. (1997) Heptology  
25:1527-1538.

Ogata, N. et al (1991) Proc. Natl. Acad. Sci. U.S.A.,  
88:3392-3396.

15

Okamoto, H., et al. (1992a). Virology 188, 331-341.

Okamoto, H., et al. (1992b) Virology 190, 894-899.

Reed, K.E., et al. (1995) J. Virol., 69:4127-4136.

20

Rice, C. M. (1996). Flaviviridae: The viruses and  
their replication, In "Fields Virology". (B. N. Fields,  
D. M. Knipe, P. M. Howley, et al., Eds.), Third ed.,  
Lippincott-Raven Publishers, Philadelphia.

Shimizu, Y.K., et al. (1990) Proc. Natl. Acad. Sci.  
U.S.A., 87:6441-6444.

25

Shimizu, Y. K., et al. (1992). Proc. Natl. Acad.  
Sci. USA 89, 5477-5481.

Shimizu, Y. K., et al. (1996). Virology 223, 409-  
412.

30

Simmonds, P., et al. (1993). J. Gen. Virol. 74,  
2391-2399.

Takamizawa, A., et al. (1991). J. Virol. 65,  
1105-1113.

35

- 70 -

Tanaka, T., et al. (1995). Biochem. Biophys. Res. Commun. 215, 744-749.

Tanaka, T., et al. (1996). Structure of the 3' terminus of the hepatitis C virus genome. J. Virol. 70, 3307-3312.

Tellier, R. et al (1996) Proc. Natl. Acad. Sci. U.S.A., 93:4370-4373.

Tellier, R., et al (1996a) J. Clin. Microbiol. 34:3085-3091.

Tsuchihara, K., et al. J. Virol. 71, 6720-6726.

Tsukiyama-Kohara, K., et al. J. Virol. 66, 1476-1483.

Weiner, A. J., et al. (1991). Virology 180, 842-848.

Weiner, A. J., et al. (1992). Proc. Natl. Acad. Sci. USA 89, 3468-3472.

Yamada, N., et al. (1996). Virology 223, 255-261.

Yanagi, M., et al. (1996). J. Infect. Dis. 174, 1324-1327.

Yanagi, M., et al. (1997). Proc. Natl. Acad. Sci. USA 94, 8738-8743.

Yoo, B. J., et al. (1995). J. Virol. 69, 32-38.



- 71 -

WHAT IS CLAIMED IS:

1. A purified and isolated nucleic acid molecule which encodes human hepatitis C virus, said molecule capable of expressing said virus when transfected into cells.

2. The nucleic acid molecule of claim 1, wherein said molecule encodes the amino acid sequence shown in Figures 14G-14H.

3. The nucleic acid molecule of claim 2, wherein said molecule comprises the nucleic acid sequence shown in Figures 14A-14F.

4. The nucleic acid molecule of claim 1, wherein said molecule encodes the amino acid sequence shown in Figures 4G-4H.

5. The nucleic acid molecule of claim 4, wherein said molecule comprises the nucleic acid sequence shown in Figures 4A-4F.

6. The nucleic acid molecule of claim 1, wherein a fragment of said molecule which encodes the structural region of hepatitis C virus has been replaced by the structural region from the genome of another hepatitis C virus strain.

7. The nucleic acid molecule of claim 6, wherein said molecule encodes the amino acid sequence shown in Figures 16G-16H.

8. The nucleic acid molecule of claim 7, wherein said molecule comprises the nucleic acid sequence shown in Figures 16A-16F.

- 72 -

°

9. The nucleic acid molecule of claim 1, wherein a fragment of the nucleic acid molecule which encodes at least one HCV protein has been replaced by a fragment of the genome of another hepatitis C virus strain which encodes the corresponding protein.

5

10. The nucleic acid molecule of claim 9, wherein the protein is selected from the group consisting of E1, E2 or NS4 proteins.

10

11. The nucleic acid molecule of claim 1, wherein a fragment of the molecule encoding all or part of an HCV protein has been deleted.

15

12. The nucleic acid molecule of claim 11, wherein the HCV protein is selected from the group consisting of P7, NS4B or NS5A proteins.

13. A DNA construct comprising a nucleic acid molecule according to claims 1, 3, 5 or 8.

20

14. An RNA transcript of the DNA construct of claim 13.

15. A cell transfected with the DNA construct of claim 13.

25

16. A cell transfected with RNA transcript of claim 14.

17. A hepatitis C virus polypeptide produced by the cell of claim 15.

18. A hepatitis C virus polypeptide produced by the cell of claim 16.

30

19. A hepatitis C virus produced by the cell of claim 13.

20. A hepatitis C virus produced by the cell of claim 14.

35

- 73 -

21. A hepatitis C virus whose genome comprises a nucleic acid molecule according to claims 1, 3, 5, 6, 8, or 9.

22. A method for producing a hepatitis C virus comprising transfecting a host cell with the RNA transcript of claim 14.

23. A polypeptide encoded by a nucleic acid sequence according to claims 1, 2, 4 or 7 or a fragment thereof.

24. The polypeptide of claim 23, wherein said polypeptide is selected from the group consisting of NS3 protease, E1 protein, E2 protein or NS4 protein.

25. A method for assaying candidate antiviral agents for activity against HCV, comprising

a) exposing a cell containing the hepatitis C virus of claim 21 to the candidate antiviral agent; and

b) measuring the presence or absence of hepatitis C virus replication in the cell of step (a).

26. The method of claim 25, wherein said replication in step (b) is measured by at least one of the following: negative strand RT-PCR, quantitative RT-PCR, Western blot, immunofluoresence, or infectivity in a susceptible animal.

27. A method for assaying candidate antiviral agents for activity against HCV, comprising:

a) exposing an HCV protease encoded by a nucleic acid sequence according to claims 1, 2, 4, or 7, or a fragment thereof to the

- 74 -

candidate antiviral agent in the presence of a protease substrate; and  
b) measuring the protease activity of said protease.

5

28. The method of claim 27, wherein said HCV protease is selected from the group consisting of an NS3 domain protease, an NS3-NS4A fusion polypeptide, or an NS2-NS3 protease.

10

29. An antiviral agent identified as having antiviral activity for HCV by the method of claim 25.

30. An antiviral agent identified as having antiviral activity for HCV by the method of claim 27.

15

31. Antibody to the polypeptide of claim 23.

32. Antibody to the hepatitis C virus of claim 21.

20

33. A method for determining the susceptibility of cells *in vitro* to support HCV infection, comprising the steps of:

a. growing animal cells in vitro;

25

b. transfecting into said cells the nucleic acid of claim 1; and

c. determining if said cells show indicia of HCV replication.

34. The method according to claim 33, wherein said cells are human cells.

30

35. A cassette vector for cloning viral genomes, comprising, inserted therein, the nucleic acid sequence according to claim 2, said vector reading in the

35

- 75 -

correct phase for the expression of said inserted sequence and having an active promoter sequence upstream thereof.

36. The cassette vector of claim 35, wherein the cassette vector is produced from plasmid pCV.

37. The cassette vector of claim 35, wherein the vector also contains one or more expressible marker genes.

38. The cassette vector of claim 35, wherein the inserted DNA sequence contains at least one ORF of the HCV genome from any strain.

39. The cassette vector of claim 35, wherein the promoter is a bacterial promoter.

40. A composition comprising a polypeptide of claim 23 suspended in a suitable amount of a pharmaceutically acceptable diluent or excipient.

41. A method for treating hepatitis C viral infection comprising the administration to a animal in need thereof of a clinically effective amount of the composition of claim 40.

42. A composition comprising a nucleic acid molecule of claim 1 suspended in a suitable amount of a pharmaceutically acceptable diluent or excipient.

43. A method for treating hepatitis C viral infection comprising the administration to an animal in need thereof of a clinically effective amount of the composition of claim 42.

1/49

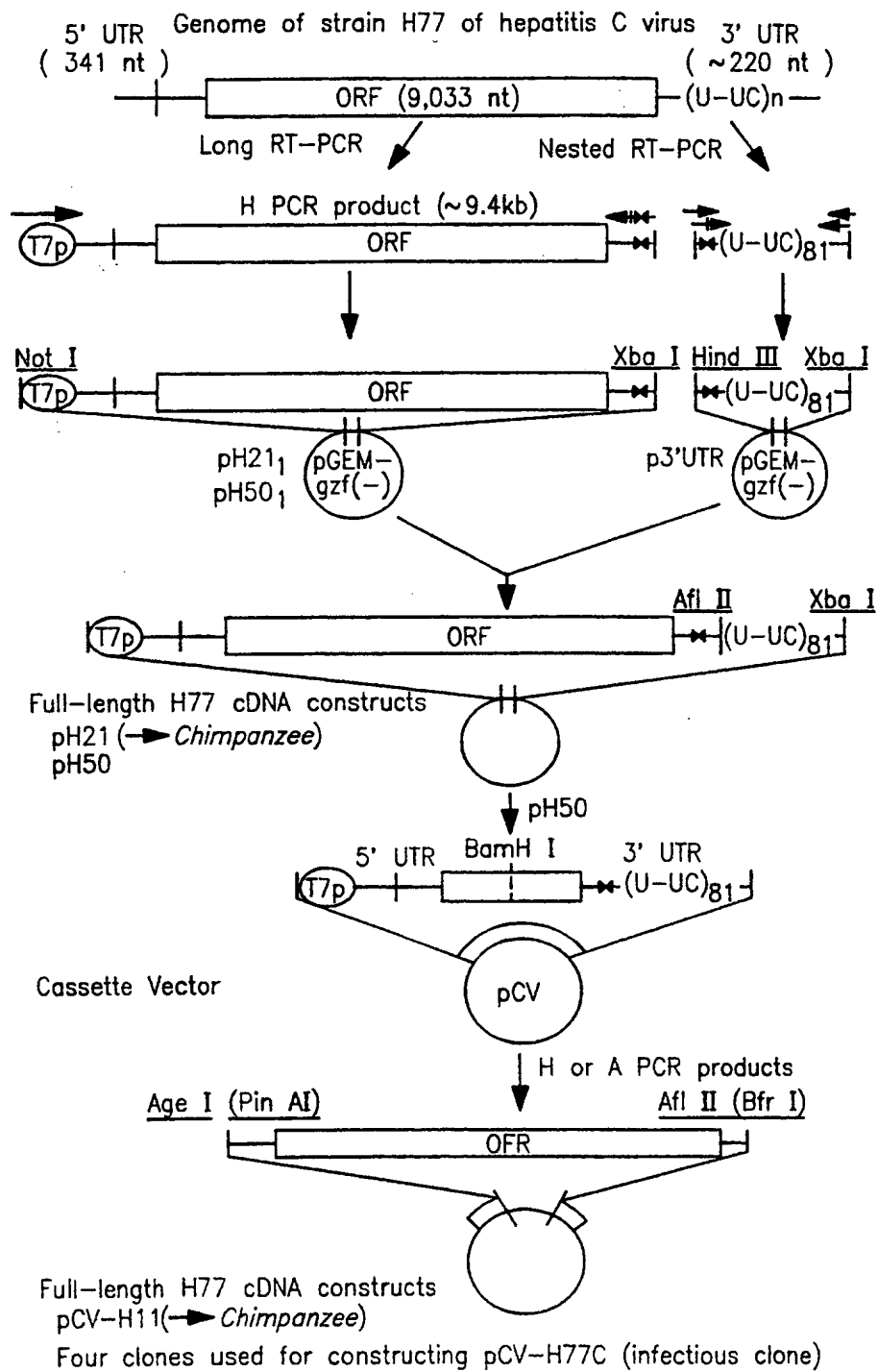


FIG. 1

SUBSTITUTE SHEET (RULE 26)

2/49

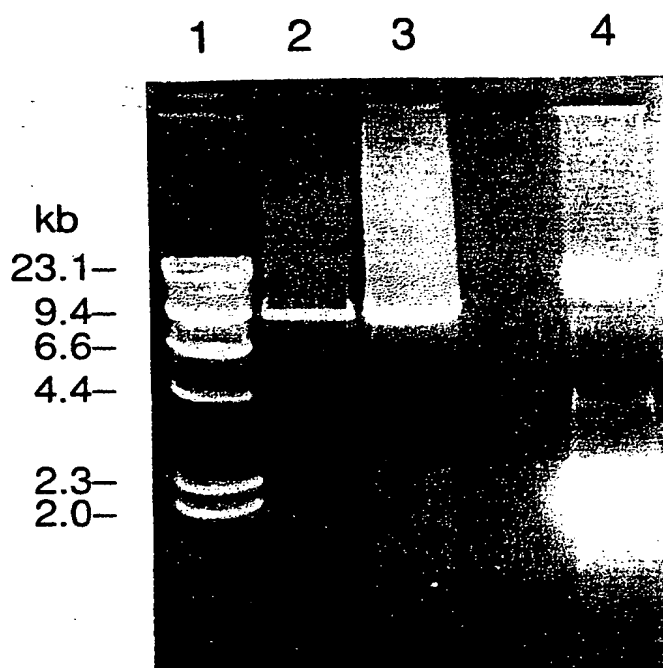


FIG. 2

3/49

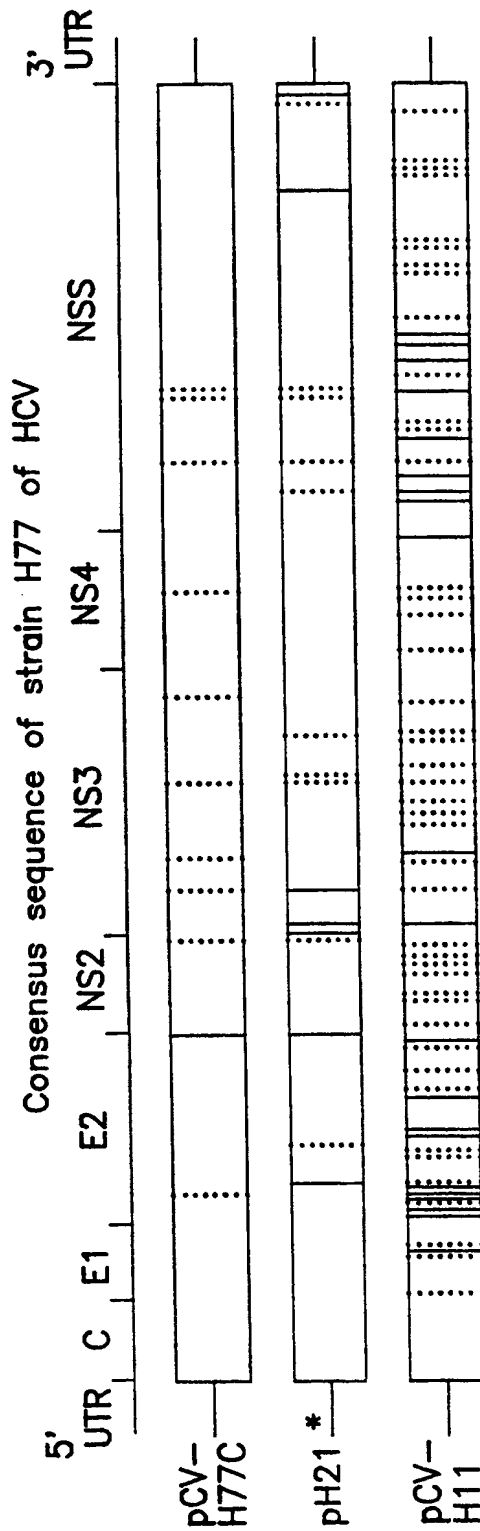


FIG. 3



4/49

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCCAGCCCCC	TGATGGGGGC	GACACTCCAC	CATGAATCAC	TCCCCGTGTA	50
GGAACACTG	TCITCAGCA	GAAAGOGTCT	AGCCATGGCG	TTAGTATGAG	100
TGTCTGTCAG	CCTCCAGGAC	CCCCCTCCC	GGGAGAGCCA	TAGTGGTCTG	150
CGGAACCGGT	GAGTACACCG	GAATTGCCAG	GACGACCGGG	TCCTTTCTTG	200
GATAAACCCG	CTCAATGCTT	GGAGATTGCG	CGGTGCCCCC	GCAAGACTGC	250
TAGCCGAGTA	GTGTGCGGTC	GCGAAAGGCC	TTGTGCTACT	GCTGATAGG	300
GTGCTTGCGA	GTGCCCCGGG	AGGCTCTGTA	GACCGTGCAC	CATGAGCAAG	350
AATCCTAAAC	CTCAAAGAAA	AACCAAACGT	AACACCAACC	GTGCCCCACA	400
GGACGTCAAG	TTCCCGGGTG	GCGGTACAGT	CGTTGGTGGG	GTCTACTTGT	450
TGCCGCGCAG	GGGCCCTAGA	TTGGGTGTGC	GCGCGACGAG	GAAGACTTCC	500
GAGCGGTCCG	AACCTCGAGG	TAGAAGTCAG	CCTATCCCCA	AGGCACGTGG	550
GCCCGAGGGC	AGGACCTGGG	CTCAGCCCCG	GTACCCCTTG	CCCCCTCTATG	600
GCAATGAGGG	TTGCCGGTGG	GCGGGATGGC	TCCTGTCTCC	CCGTGGCTCT	650
CGGCCTAGCT	GGGGCCCCAC	AGACCCCCCG	CGTAGGTCCG	GCAATTGGGG	700
TAAGGTATC	GATACCCCTA	CGTGCGGCTT	CGCCGACCTC	ATGGGGTACA	750
TACCGCTCGT	CGGCGCCCTT	CTTGGAGGGG	CTGCCAGGGC	CCTGGCGCAT	800
GGCGTCCGGG	TTCTGGAAGA	CGGCGTGAAC	TATGCAACAG	GGAACCTTCC	850
TGGTTGCTCT	TTCTCTATCT	TCCTTCTGGC	CCTGCTCTCT	TGCCCTGACTG	900
TGCCCCGCTT	AGCCTACCAA	GTGCGCAATT	CCTCGGGGCT	TTACCATGTC	950
ACCAATGATT	GCCCTAACTC	GAGTATTGTG	TACGAGGCGG	CCGATGCCAT	1000
CCTGCACACT	CCGGGGTGTG	TCCCTTGCGT	TGCGGAGGGT	AACGCCCTCGA	1050
GGTGTGTGGT	GGCGGTGACC	CCACGGTGGG	CCACCAGGGA	CGGCAAACCTC	1100
CCCACAACGC	AGCTTCGACG	TCATATCGAT	CTGCTTGTGG	GGAGCGCCAC	1150
CCTCTGCTCG	GCCCTCTACG	TGGGGGACCT	GTGCGGGTCT	GTCTTTCTTG	1200
TTGGTCAACT	GTCTACCTTC	TCTCCCAGGC	GCCACTGGAC	GACGCAAGAC	1250
TGCAATTGTT	CTATCTATCC	CGGCCATATA	ACGGGTATATC	GCATGGCATG	1300
GGATATGATG	ATGAACCTGG	CCCCTACGGC	AGGGTTGGTG	GTAGCTCAGC	1350
TGCTCCGGAT	CCCACAAGCC	ATCATGGACA	TGATCGCTGG	TGCTCACTGG	1400
GGAGTCCCTG	CGGGCATAGC	GTATTCTCTC	ATGGTGGGGA	ACTGGGCGAA	1450
GGTCTGGTGA	GTGCTGCTGC	TATTTGCCGG	CGTCCGACGG	GAAACCCACG	1500
TCACCGGGGG	AAATGCCCGC	CGCACCACGG	CTGGGCTTGT	TGGTCTCCTT	1550
ACACCAGGCG	CCAAGCAGAA	CATCCAACCTG	ATCAACACCA	ACGGCAGTTG	1600
GCACATCAAT	AGCACGGCCT	TGAATTGCAA	TGAAAGCCTT	AACACCGGCT	1650
GGTTAGCAGG	GCTCTTCTAT	CAACACAAAT	TCAACTCTTC	AGGCTGTCTT	1700
GAGAGGTGCG	CCAGCTGCGG	ACGCCCTTACC	GATTTTGCCC	AGGGCTGGGG	1750
TCCTATCAGT	TATGCCAAGG	GAAGCGGCCT	CGACCAACGC	CCCTACTGCT	1800
GGCACTACCC	TCCAAGACCT	TGTGGCATTG	TGCCCCGAAA	GAGCGTGTGT	1850
GGCCCGGTAT	ATTGCTTCAC	TCCCAGCCCC	GTGGTGGTGG	GAACGACCGA	1900

FIG. 4A

SUBSTITUTE SHEET (RULE 26)

## H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAGGTGGGGC	GCGCTACCT	ACAGCTGGGG	TGCAAATGAT	ACGGATGTCT	1950
TGTCCTTAA	CAACACCAGG	CCACCGCTGG	GCAATTGGTT	CGGTGTGACC	2000
TGGATGAACT	CAACTGGATT	CACCAAAGTG	TGCGGAGGCG	CCCTTGTGT	2050
CATCGGAGGG	GTGGGCAACA	ACACCTTGCT	CTGCCCCACT	GATTGCTTCC	2100
GCAAACATCC	GGAAGCCACA	TACTCTGGGT	GCGGCTCCGG	TCCCTGGATT	2150
ACACCCAGGT	GCATGGTCCA	CTACCCGTAT	AGGCCTTGGC	ACTATCCTTG	2200
TACCATCAAT	TACACCATAT	TCAAAGTCAG	GATGTACGTG	GGAGGGGTGG	2250
AGCACAGGCT	GGAAGCGGCC	TGCAACTGGA	CGCGGGGGGA	AAGCTGTGAT	2300
CTGGAAGACA	GGGACAGGTC	CGAGCTCAGC	CGGTGTCTGC	TGTCCACCAC	2350
ACAGTGGCAG	GTCTTTCGGT	GTCTTTTCAC	GACCCGTCCA	GCCTTGTCCA	2400
CCGGCCCTCAT	CCACCTCCAC	CAGAACATTG	TGGACGTGCA	GTAATTGTAC	2450
GGGGTAGGGT	CAAGCATCGC	GTCTTGGGCC	ATTAAAGTGGG	AGTACGTCTG	2500
TCTCCTGTTC	CTTCTGTCTG	CAGACGGCGG	CGTCTGCTCC	TGCTTGTGGA	2550
TGATGTTACT	CATATCCCAA	GCGGAGGCGG	CTTTGGAGAA	CCCTGTAATA	2600
CTCAATGCAG	CATCCCTGGC	CGGGACGCAC	GGTCTTGTGT	CCCTCCTCGT	2650
GTCTTCTGTC	TTTGGCGTGT	ATCTGAAGGG	TAGGTGGGTG	CCCGGAGCGG	2700
TCTACGCCCT	CTACGGGATG	TGGCCTCTCC	TCCCTGCTCCT	GCTGGCGTGT	2750
CCTCAGCGGG	CATACGCACT	GGACACGGAG	GTGGCCGCGT	CGTGTGGCGG	2800
CGTGTGTCTT	GTGGGGTTAA	TGGCGCTGAC	TCTGTGCGCA	TATTACAAGC	2850
GCTATATCAG	CTGGTGCATG	TGGTGGCTTC	AGTATTTTCT	GACCAGAGTA	2900
GAAGCGCAAC	TGCACGTGTG	GGTTCCCCCC	CTCAACGTCC	GGGGGGGGCG	2950
CGATGCCGTC	ATCTTACTCA	TGTGTGTAGT	ACACCCGACC	CTGGTATTTG	3000
ACATCACCAA	ACTACTCCTG	GCCATCTTCG	GACCCCTTTG	GATTCTTCAA	3050
GCCAGTTTGC	TTAAAGTCCC	CTACTTCGTG	CGGTTCAAG	GCCTTCTCCG	3100
GATCTGGCGG	CTAGCGCGGA	AGATAGCCGG	AGGTCAATTAC	GTGCAAATGG	3150
CCATCATCAA	GTTAGGGGGG	CTTACTGGCA	CCTATGTGTA	TAACCATCTC	3200
ACCCCTCTTC	GAGACTGGGC	GCACAACGGC	CTGGGAGATC	TGGCCGTGGC	3250
TGTGGAACCA	GTGGTCTTCT	CCCGAATGGA	GACCAAGCTC	ATCAOGTGGG	3300
GGGCAGATAC	CGCCGCGTGC	GGTGACATCA	TCAACGGCTT	GCCCGTCTCT	3350
GCCCCGTAGGG	GCCAGGAGAT	ACTGCTTGGG	CCAGCCGACG	GAATGGTCTC	3400
CAAGGGGTGG	AGGTGTCTGG	CGCCCATCAC	GGCGTACGCC	CAGCAGACGA	3450
GAGGCCCTCCT	AGGGTGTATA	ATCACCAGCC	TGACTGGCCG	GCACAAAAC	3500
CAAGTGGAGG	GTGAGGTCCA	GATCGTGTCA	ACTGCTACCC	AAACCTTCCT	3550
GGCAACGTGC	ATCAATGGGG	TATGCTGGAC	TGTCTACCAC	GGGGCCGGAA	3600
CGAGGACCAT	CGCATCACCC	AAGGGTCCCT	TCATCCAGAT	GTATACCAAT	3650
GTGGACCAAG	ACCTTGTGGG	CTGGCCCGCT	CCTCAAGGTT	CCCGCTCATT	3700
GACACCCCTGT	ACCTGCGGCT	CCTCGGACCT	TTACCTGGTC	ACGAGGCACG	3750
CCGATGTICAT	TCCCGTGGCG	CGGCGAGGTG	ATAGCAGGGG	TAGCCTGCTT	3800

FIG. 4B

SUBSTITUTE SHEET (RULE 26)

6/49

## H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGCCCCCGGC	CCATTTCCTA	CTTGAAAGGC	TCCTCGGGGG	GTCGCTGTT	3850
GTGCCCCGCG	GGACACGCG	TGGGCTATT	CAGGGGCGCG	GTGTGCAACC	3900
GTGGAGTGGC	TAAAGCGGTG	GACTTTATCC	CTGTGGAGAA	CCTAGGGACA	3950
ACCATGAGAT	CCCCGGTGT	CACGGACAAC	TCCTCTCCAC	CAGCAGTGGC	4000
CCAGAGCTTC	CAGGTGGGCC	ACCTGCATGC	TCCCACCGGC	AGCGGTAGA	4050
GCACCAAGGT	CCCGGCTGG	TACGCAGGCC	AGGGCTACAA	GGTGTGGTG	4100
CTCAACCCCT	CTGTGTCTGC	AACGCTGGGC	TTTGGTGTCT	ACATGTCCAA	4150
GGCCCATGGG	GTGTATCCTA	ATATCAGGAC	CGGGGTGAGA	ACAATTACCA	4200
CTGGCAGCCC	CATCACGTAC	TCACCTACG	GCAAGTTCCT	TGCGCAGGC	4250
GGGTGCTCAG	GAGGTGCTTA	TGACATAATA	ATTTGTGACG	AGTGCCACTC	4300
CACGGATGCC	ACATCCATCT	TGGGCATCGG	CAGTGTCTCT	GACCAAGCAG	4350
AGACTGGCGG	GGCGAGACTG	GTGTGTCTCG	CCACTGCTAC	CCCTCCGGGC	4400
TCCGTCACTG	TGTCCCATCC	TAACATCGAG	GAGGTGTCTC	TGTCCACCAC	4450
CGGAGAGATC	CCCTTTTACG	GCAAGGCTAT	CCCCCTCGAG	GTGATCAAGG	4500
GGGGAAGACA	TCTCATCTTC	TGCCACTCAA	AGAAGAAGTG	CGACGAGCTC	4550
GCCGCGAAGC	TGGTCCGATT	GGGCATCAAT	GCCGTGGCCT	ACTACCGCGG	4600
TCTTGACGTG	TCTGTATCC	CGACCAGCGG	CGATGTGTCT	GTGCTGTCTG	4650
CCGATGCTCT	CATGACTGGC	TTTACCGGGG	ACTTGGACTC	TGTGATAGAC	4700
TGCAACACGT	GTGTCACTCA	GACAGTCGAT	TTCAGCCTTG	ACCCATACCT	4750
TACCATTGAG	ACAACCACGC	TCCCCCAGGA	TGCTGTCTCC	AGGACTCAAC	4800
GCCGGGGCAG	GACTGGCAGG	GGGAAGCCAG	GCATCTATAG	ATTTGTGGCA	4850
CCGGGGGAGC	GCCCCCTCGG	CATGTTCGAC	TGCTCCGTCC	TCTGTGAGTG	4900
CTATGACCGG	GGCTGTGCTT	GGTATGAGCT	CACGCCCCGC	GAGACTACAG	4950
TTAGGCTACG	AGCGTACATG	AACACCCCGG	GGCTTCCCGT	GTGCCAGGAC	5000
CATCTTGAAT	TTTGGGAGGG	CGTCTTTACG	GGCTCACTC	ATATAGATGC	5050
CCACTTTTTTA	TCCCAGACAA	AGCAGAGTGG	GGAGAACTTT	CCTTACCTGG	5100
TAGCGTACCA	AGCCACCGTG	TGCGCTAGGG	CTCAAGCCCC	TCCCCCATCG	5150
TGGGACCAGA	TGTGGAAGTG	TTTGATCCGC	CTTAAACCCA	CCCTCCATGG	5200
GCCAACACCC	CTGCTATACA	GACTGGGCGC	TGTTTCAGAAT	GAAGTCAACC	5250
TGACGCACCC	AATCACCAAA	TACATCATGA	CATGCATGTC	GGCCGACCTG	5300
GAGGTCTGTA	CGAGCACCTG	GGTGTCTGTT	GGCGCGTCC	TGGCTGCTCT	5350
GGCCGCGTAT	TGCTGTCTAA	CAGGCTGGGT	GGTCATAGTG	GGCAGGATCG	5400
TCTTGTCCGG	GAAGCCGGCA	ATTATACCTG	ACAGGGAGGT	TCTCTACCAG	5450
GAGTTCGATG	AGATGGAAGA	GTGCTCTCAG	CAGTTACCGT	ACATCGAGCA	5500
AGGGATGATG	CTCGCTGAGC	AGTTCAAGCA	GAAGGCCCTC	GGCTCTCTGC	5550
AGACCGCGTC	CCGCCATGCA	GAGGTATATCA	CCCTGCTGT	CCAGACCAAC	5600
TGGCAGAAAC	TCCAGGTCTT	TTGGGCGAAG	CACATGTGGA	ATTTTCATCAG	5650
TGGCATACAA	TACTTGGCGG	GCCTGTCAAC	GCTGCCTGGT	AACCCCGCCA	5700

FIG. 4C

7/49

## H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TTGCTTCATT	GATGGCTTTT	ACAGCTGCCG	TCACCAGCCC	ACTAACCCT	5750
GGCCAAACCC	TCCTCTTCAA	CATATTTGGG	GGGTGGGIGG	CTGCCCAGCT	5800
CGCCGCCCCC	GGTGGCGCTA	CTGCTTTTGT	GGGTGCTGGC	CTAGCTGGCG	5850
CCGCCATCCG	CAGCGTTGGA	CTGGGGAAGG	TCCTCGTGGG	CATTCTTGCA	5900
GGGTATGGCG	CGGGGCTGGC	GGGAGCTCTT	GTAGCATTCG	AGATCATGAG	5950
CGGTGAGGTC	CCCTCCACGG	AGGACCTGGT	CAATCTGCTG	CCCGCATCC	6000
TCTCGCTGG	AGCCTTTGTA	GTGGGTGGG	TCTGGCGAGC	AATACTGGCG	6050
CGGCACGTTG	GCCCCGGCGA	GGGGGCAGTG	CAATGGATGA	ACCGGCTAAT	6100
AGCCTTCGCC	TCCCGGGGGA	ACCATGTTTC	CCCCACGCAC	TACGTGCCGG	6150
AGAGCGATGC	AGCCGCCCCG	GTCACCTGCC	TACTCAGCAG	CCTCAGTGTG	6200
ACCCAGCTCC	TGAGGCGACT	GCATCAGTGG	ATAAGCTCGG	AGTGTACAC	6250
TCCATGCTCC	GGTTCCTGGC	TAAGGGACAT	CTGGGACTGG	ATATGCGAGG	6300
TGCTGAGCGA	CTTTAAGACC	TGGCTGAAAG	CCAAGCTCAT	GCCACAACCTG	6350
CCTGGGATTC	CCTTTGTGTC	CTGCCAGCGC	GGGTATAGGG	GGGTCTGGCG	6400
AGGAGACGGC	ATTATGCACA	CTCGCTGCCA	CTGTGGAGCT	GAGATCACTG	6900
GACATGTCAA	AAACGGGACG	ATGAGGATCG	TGGTTCCTAG	GACCTGCAGG	6950
AACATGTGGA	GTTGGACGTT	CCCCATTAAC	GCTTACACCA	CGGGCCCCCTG	6550
TACTCCCCCT	CCTGCGCCGA	ACTATAAGTT	CGCGCTGTGG	AGGGTGTCTG	6600
CAGAGGAATA	CGTGGAGATA	AGCGGGGTGG	GGGACTTCCA	CTACGTATCG	6650
GGTATGACTA	CTGACAATCT	TAAATGCCCG	TGCCAGATCC	CATCGCCCCG	6700
ATTTTTCACA	GAATTGGACG	GGGTGCGCCT	ACACAGGTTT	GCGCCCCCTT	6750
GCAAGCCCTT	GCTGCGGGAG	GAGGTATCAT	TCAGAGTAGG	ACTCCACGAG	6800
TACCCGGTGG	GGTGGCAATT	ACCTTGCGAG	CCCCAACCCG	ACGTAGCCGT	6850
GTGTACGTCC	ATGCTCACTG	ATCCCTCCCA	TATAACAGCA	GAGGCGGCGG	6900
GGAGAAGGTT	GGCGAGAGGG	TCACCCCTTT	CTATGGCCAG	CTCCTCGGCT	6950
AGCCAGCTGT	CCGCTCCATC	TCTCAAGGCA	ACTTGCACCG	CCAACCATGA	7000
CTCCCCCTGAC	GCCGAGCTCA	TAGAGGCTAA	CCTCCTGTGG	AGGCAGGAGA	7050
TGGGCGGCAA	CATCACCAGG	GTGTAGTCAG	AGAACAAGT	GGTGATTCTG	7100
GACTCCCTTCG	ATCCGCTTGT	GGCAGAGGAG	GATGAGCGGG	AGGTCTCCGT	7150
ACCTGCAGAA	ATTCTGCGGA	AGTCTCGGAG	ATTGCCCCGG	GCCCTGCCCG	7200
TCTGGGCGCG	GCCGGACTAC	AACCCCCCGC	TAGTAGAGAC	GTGGAAAAAG	7250
CCTGACTACG	AACCACCTGT	GGTCCATGGC	TGCCCCCTAC	CACCTCCACG	7300
GTCCCCCTCCT	GTGCCCTCCG	CTCGGAAAAA	GCGTACGGTG	GTCTCAACCG	7350
AATCAACCCCT	ATCTACTGCC	TTGGCCGAGC	TTGCCACCAA	AAGTTTITGGC	7400
AGCTCCTCAA	CTTCCGGCAT	TACGGGCGAC	AATACGACAA	CATCCTCTGA	7450
GCCCCGCCCT	TCTGGCTGCC	CCCCCGACTC	CGAGGTGAG	TCCTATTCTT	7500
CCATGCCCCC	CCTGGAGGGG	GAGCCTGGGG	ATCCGGATCT	CAGCGACGGG	7550
TCATGGTCCA	CGGTCACTAG	TGGGGCCGAC	ACGGAAGATG	TCGTGTGCTG	7600

FIG. 4D

SUBSTITUTE SHEET (RULE 26)

8/49

## H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CTCAATGICT	TATTCCTGGA	CAGGCGCACT	CGTACCCCCG	TGCGCTGCGG	7650
AAGAACA AAA	ACTGCCCCATC	AACGCACTGA	GCAACTCGTT	GCTACGCCAT	7700
CACAATCTGG	TGTATTCCAC	CACTTCACGC	AGTGCTTGCC	AAAGGCAGAA	7750
GAAAGTCACA	TTTGACAGAC	TGCAAGTTCT	GGACAGCCAT	TACCAGGACG	7800
TGCTCAAGGA	GGTCAAAGCA	GCGGCGTCAA	AAGTGAAGGC	TAACITGCTA	7850
TOCGTAGAGG	AAGCTTGCAG	CCTGACGCCC	CCACATTGAG	CCAAATCCAA	7900
GTTTGGCTAT	GGGGCAAAAG	ACGTCCGTTC	CCATCCCGAG	AAGGCCGTAG	7950
CCCACATCAA	CTCCGTGTGG	AAAGACCTTC	TGGAAGACAG	TGTAACACCA	8000
ATAGACACTA	CCATCATGGC	CAAGAACGAG	GTTCCTGCGG	TTCAGCCTGA	8050
GAAGGGGGGT	CGTAAGCCAG	CTCGTCTCAT	CGTGTTCGCC	GACCTGGGCG	8100
TGCGCGTGTG	CGAGAAGATG	GCCCTGTACG	ACGTGGTTAG	CAAGCTCCCC	8150
CTGGCCGTGA	TGGGAAGCTC	CTACGGATTG	CAATACTCAC	CAGGACAGCG	8200
GGTGAATTC	CTCGTGAAG	CGTGAAGTTC	CAAGAAGACC	CCGATGGGGT	8250
TCTCGTATGA	TACCCGCTGT	TTTGACTCCA	CAGTCACTGA	GAGCGACATC	8300
CGTACGGAGG	AGGCAATTGA	CCAATGTTGT	GACCTGGACC	CCCAAGCCCC	8350
CGTGGCCATC	AAGTCCCTCA	CTGAGAGGCT	TTATGTTGGG	GGCCCTCTTA	8400
CCAATTCAAG	GGGGGAAAC	TGCGGCTACC	GCAGGTGCGG	CGCGAGCGGC	8450
GTA CTGACAA	CTAGCTGTGG	TAACACCCCTC	ACTTGCTACA	TCAAGGCCCG	8500
GGCAGCCTGT	CGAGCCGCAG	GGCTCCAGGA	CTGCACCATG	CTCGTGTGTG	8550
GCGACGACTT	AGTCGTATATC	TGTGAAAGTG	CGGGGGTCCA	GGAGGACGCG	8600
GCGAGCCTGA	GAGCCTTCAC	GGAGGCTATG	ACCAGGTACT	CCGCCCCCCC	8650
CGGGGACCCC	CCACAACCAG	AATACGACTT	GGAGCTTATA	ACATCATGCT	8700
CCTCCAACGT	GTCAGTCCGC	CACGACGGCG	CTGGAAAGAG	GGTCTACTAC	8750
CTTACCCGTG	ACCCTACAAC	CCCCCTCGCG	AGAGCCGCGT	GGGAGACAGC	8800
AAGACACACT	CCAGTCAATT	CCTGGCTAGG	CAACATAATC	ATGTTTGGCC	8850
CCACACTGTG	GGCGAGGATG	ATACTGATGA	CCCATTTCTT	TAGCGTCTTC	8900
ATAGCCAGGG	ATCAGCTTGA	ACAGGCTCTT	AACTGTGAGA	TCTACGGAGC	8950
CTGCTACTCC	ATAGAACCAC	TGGATCTACC	TCCAATCATT	CAAAGACTCC	9000
ATGGCCTCAG	CGCATTTTCA	CTCCACAGTT	ACTCTCCAGG	TGAAATCAAT	9050
AGGGTGGCCG	CATGCCTCAG	AAAACCTGGG	GTCCCGCCCT	TGCGAGCTTG	9100
GAGACACCCG	GCCCCGAGCG	TCCGCGCTAG	GCTTCTGTCC	AGAGGAGGCA	9150
GGGCTGCCAT	ATGTGGCAAG	TACCTCTTCA	ACTGGGCAGT	AAGAACAAG	9200
CTCAA ACTCA	CTCCAATAGC	GGCCGCTGGC	CGGCTGGACT	TGTCCGGTTG	9250
GTTACAGGCT	GGCTACAGCG	GGGGAGACAT	TTATCACAGC	GTGTCTCATG	9300
CCCGGCCCCG	CTGGTCTGTG	TTTTCCTTAC	TCTTGTCTGC	TGCAGGGGTA	9350
GGCATCTACC	TCTTCCCCAA	CCGATGAAGG	TTGGGGTAAA	CACTCCGGCC	9400
TCTTAAGCCA	TTTCCGTGTT	TTTTTTTTTT	TTTTTTTTTT	TTTTTCTTTT	9450
TTTTTTTCTT	TCTTTTCTTT	CTTTTTTTTC	TTTCTTTTTT	CCTTCTTTAA	9500

FIG. 4E

SUBSTITUTE SHEET (RULE 26)

9/49

H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGGTGGCTCC	ATCTTAGCCC	TAGTCACGGC	TAGCTGIGAA	AGGTCCGIGA	9550
GCCGCATGAC	TGCAGAGAGT	GCTGATACTG	GCCTCTCTGC	AGATCATGT	9599

FIG. 4F

10/49

## H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
MSINPKPQRK	TKRNINRRPQ	DVKFPGGGQI	VGGVYLLPRR	GPRLGVRATR	50
KTSESRQPRG	RRQPIPKARR	PEGRIWAQFG	YFWPLYGNEG	CGWAGWLLSP	100
RGSRPSWGPT	DFRRRSRNLG	KVIDTLTQGF	ADLMGYIPLV	GAPLGGAARA	150
LAHGVRVLED	GVNYATGNLP	GCSFSIFLLA	LLSCLTVPAS	AYQVRNSSGL	200
YHVINDCPNS	SIVYEAADAI	LHTFGCVPCV	REGNASRCWV	AVTPTVATRD	250
GKLPTTQLRR	HIDLLVGSAT	LCSALYVGDL	CGSVFLVGQL	FTFSRRRHWT	300
TQDQNCSTYP	GHTTGHMAW	IMMNWSPTA	ALVVAQLLRI	PQAIMDMIAG	350
AHWGLAGIA	YFSMVGWAK	VLVWLLLFAG	VDAETHVTGG	NAGRTTAGLV	400
GLLTGPAKQN	IQLININGSW	HINSTALNQN	ESLNTGWLAG	LFYQHKFNSS	450
GCPERLASCR	RLTDEAQQWG	PISYANGSGL	DERPYOWHYP	PRFCGIVPAK	500
SVCGPVYCF	PSPVVVGTTD	RSGAPTYSWG	ANDIDVFLN	NIRPPLGNWF	550
GCTWMNSTGF	TKVOGAPPCV	IGGVGNILL	CPIDCFRKHP	EATYSRCGSG	600
FWITPRQMD	YPYRLWHYPC	TINYTIFKVR	MYVGGVEHRL	EAACNWIRGE	650
RCDLEDRLRS	ELSPLLLSTT	QWQVLPCSFT	TLPALSTGLI	HLHQNIQDVQ	700
YLYGVGSSIA	SWAIKWEYV	LLFLLADAR	VCSCIWMMLL	ISQAEAALEN	750
LVIILNAASLA	GTHGLVSFLV	FFCFAWYKLG	RWVPGAVYAL	YGMWPLLLLL	800
LALPQRAYAL	DTEVAASCGG	WLVGLMALT	LSPYKRYIS	WOMMWLQYFL	850
TRVEAQLHMW	VPPLNVRGGR	DAVILLMCV	HPTLVFDITK	LLLAIFGPLW	900
TLQASLLKVP	YFVRVQGLLR	ICALARKIAG	GHYVQMAIK	LGALTGTIVY	950
NHLTPLRDWA	HNGLRDLAVA	VEPVVFSRME	TKLITWGADT	AACGDIINGL	1000
PVSARRGQEI	LLGPADGMVS	KGWRLAPIT	AYAQOTRGLL	GCIITSLTGR	1050
DKNQVEGEVQ	IVSTATQIFL	ATCINGVCWT	VYHGAGIRTI	ASPKGFVIQM	1100
YTINVDQDLVG	WPAPQGSRL	TPCTCGSSDL	YLVTRHADVI	FVRRRGDSRG	1150
SLLSRPISY	LKGSSGGPLL	CPAGHAVGLF	RAAVCIRGVA	KAVDFIPVEN	1200
LGTIMRSPVF	TDNSSPPAVP	QSFQVAHLHA	PTGSGKSTKV	PAAYAAQGYK	1250
VLVLNPSVAA	TLGFGAYMSK	AHGVDENIRT	GVRITTTGSP	ITYSTYKFL	1300
ADGGCSCGAY	DIICDECHS	TDATSILGIG	TVLDQAETAG	ARLVLATAT	1350
PPGSVIVSHP	NIEEVALSTT	GEIPFYGKAI	PLEVIKGRH	LIFCHSKKCC	1400
DELAACKLVAL	GINAVAYYRG	LDSVIPTSG	DVVVSTIDAL	MIGFTGDFDS	1450
VIDCNTCVTQ	TVDFSLDPTF	TIETTTLPQD	AVSRTQRRGR	TGRGKPGIYR	1500
FVAPGERPSG	MFDSSVLCEC	YDAGCAWYEL	TPAETTVRLR	AYMNTFGLPV	1550
CQDHLEFWEG	VFTGLTHIDA	HFLSQTKQSG	ENFPYLVAYQ	ATVCARAQAP	1600
PPSWDQMWKC	LIRLKPTLHG	PTPLLYRLGA	VQNEVTILHP	ITKYIMTOMS	1650
ADLEVVTSTW	VLVGVLAAL	AAYCLSTGCV	VIVGRIVLSG	KPAIIPDREV	1700
LYQEFDEMEE	CSQHLEPYIEQ	GMLAEQFKQ	KALGLLQIAS	RHAEVITPAV	1750
QTNWQKLEVF	WAKHMANFIS	GIQYLAGLST	LPGNPAIASL	MAFTAAVTSP	1800
LTTGQITLLFN	ILGGWAAQL	AAPGAATAFV	GAGLAGAAIG	SVGLGKVLVD	1850
ILAGYGAGVA	GALVAFKIMS	GEVPSTEDLV	NLLPAILSPG	ALVVGVCVAA	1900

FIG. 4G

SUBSTITUTE SHEET (RULE 26)

11/49

## H77C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ILRRHVGPGE	GAVQWMNRLI	AFASRGNHVS	PIHYVPESDA	AARVTAIILSS	1950
LITVTQLLRRL	HQWISSECTT	PCSGSWLRDI	WDWICEVLSD	FKTWLKAKLM	2000
PQLPGIPFVS	CQRGYRGVWR	GDGIMHIRCH	CGAETTGHVK	NGIMRIVGPR	2050
TCRNMWSGIT	PINAYTIGPC	TPLPAPNYKF	ALWRVSAEY	VEIRRVGDFH	2100
YVSGMITDNL	KCPQIQPSPE	FFTELDGVRL	HRFAPPCKPL	LREEVSFRVG	2150
LHEYFVGSQI	PCEPEPDVAV	LTSMLTDPHS	ITAEAAGRRL	ARGSPPSMAS	2200
SSASQLSAPS	LKATCTANHD	SPDAELIEAN	LLWRQEMGGN	ITRVESENKV	2250
VILDSFDPLV	AEEDEREVS	PAETLRKSRR	FARALPWAR	PDYNPPLVET	2300
WKKPDYEPPV	VHGCPLPPPR	SPPVPPPRKK	RIVVLTESTL	STALAEIATK	2350
SFGSSSTSGI	TGDNITTSSE	PAPSGCFFDS	DVESYSSMPP	LEGERGDPDL	2400
SDGSWSIVSS	GADTEDVOC	SMSYSWIGAL	VTPCAAEEQK	LPINALSNSL	2450
LRHHNLVYST	TSRSACQROK	KVTFDRLOVL	DSHYQDLKE	VKAAASKVKA	2500
NLLSVEEACS	LTPPHSAKSK	FGYGAKDVRC	HARKAVAHIN	SWKDLLED	2550
VTPIDTTIMA	KNEVFCVQPE	KGGRKPARLI	VFPDLGVRC	EKMALYDVS	2600
KLPLAVMGSS	YGFOYSPGQR	VEFLVQAWKS	KKTPMGFSYD	TRCFDSTVTE	2650
SDIRTEEATY	QCCDLDPQAR	VAIKSLTERL	YVGGPLINSR	GENCGYRRCR	2700
ASGVLTTSCG	NLTICYIKAR	AACRAAGLQD	CTMLVCGDDL	VVICESAGVQ	2750
EDAASLRAFT	EAMTRYSAAP	GDPPQPEYDL	ELITSCSSNV	SVAHDGAGKR	2800
VYYLTRDPTT	PLARAAWETA	RHTPVNSWLG	NIIMFAPTLW	ARMILMIHFF	2850
SVLIARDQLE	QALNCETYGA	CYSIEPLDLP	PIIQRLHGLS	AFSLHSYSFG	2900
EINRVAACLR	KLGVPPPLRAW	RHRARSVRAR	LLSRGGRAAI	CGKYLEFNWAV	2950
RTKLKLTPIA	AAGRDLDSGW	FTAGYSGGDI	YHSVSHARPR	WFWFCLLLLA	3000
AGVGITYLLPN	R				3011

FIG. 4H



12/49

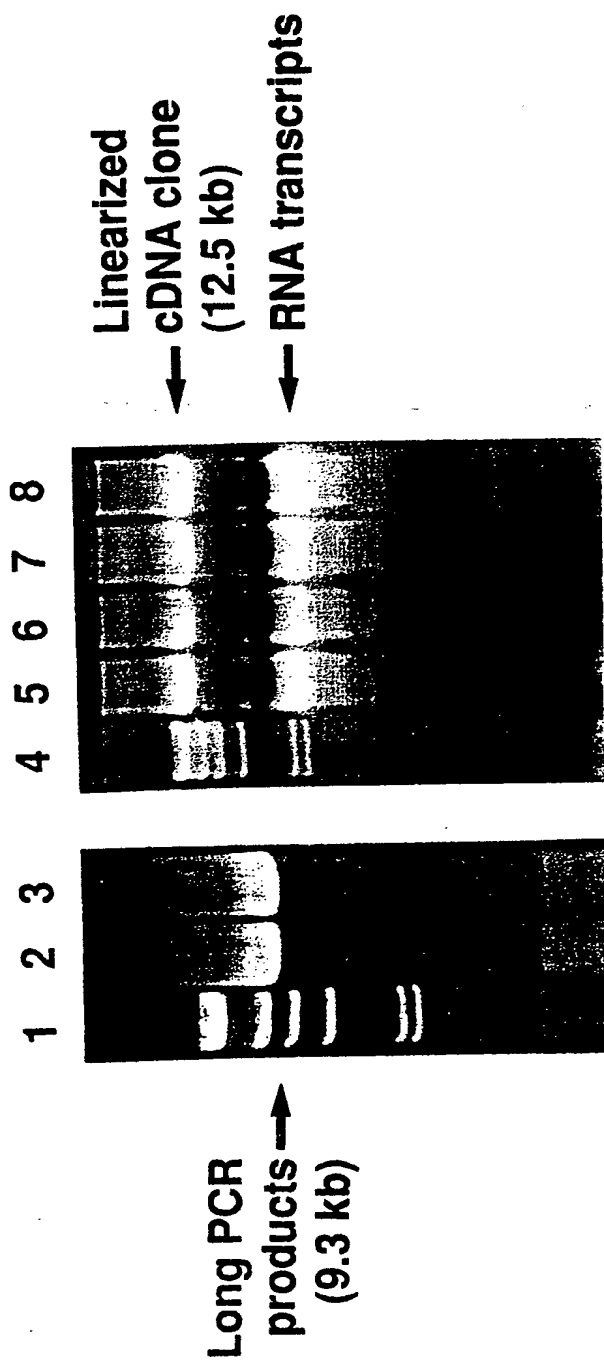
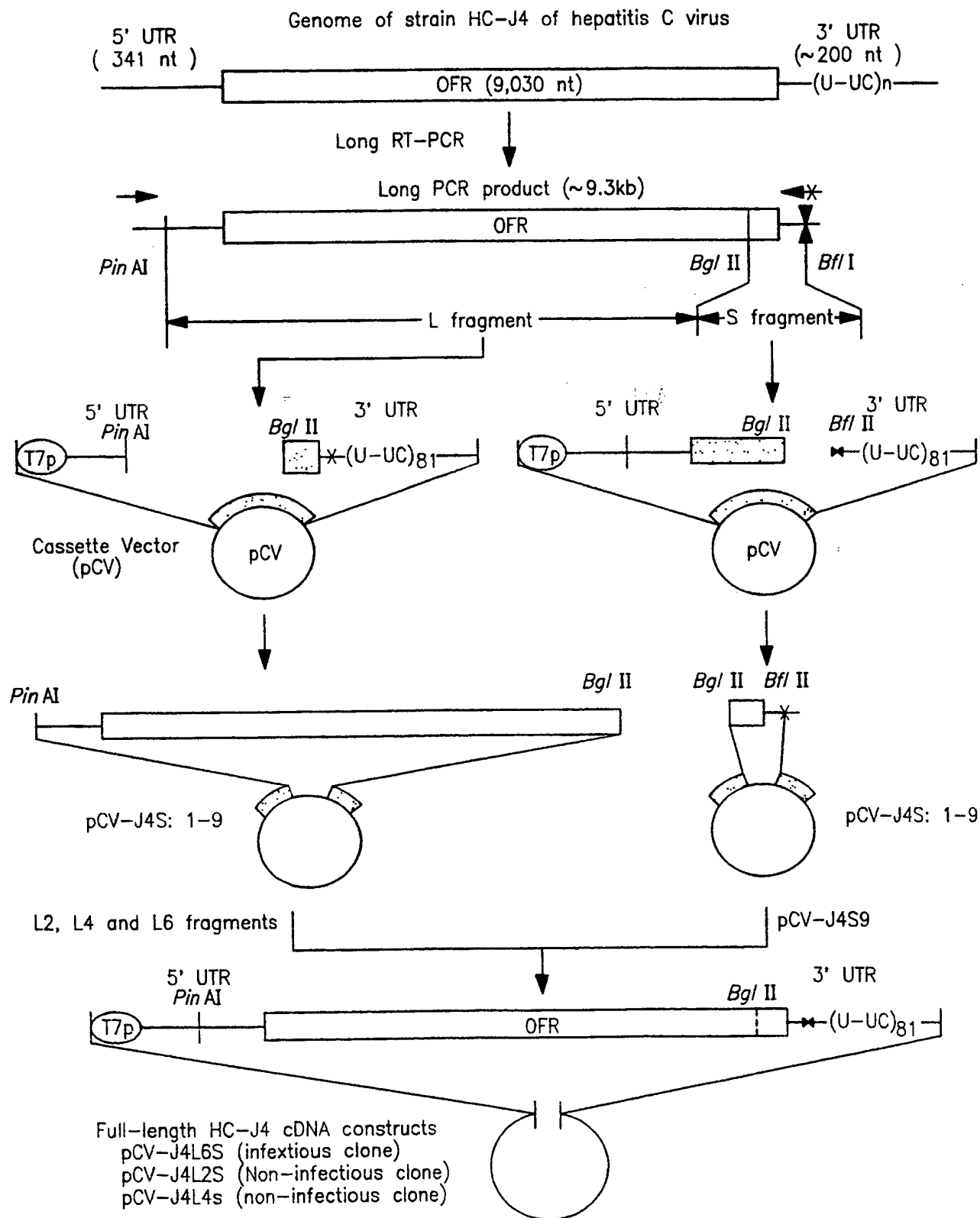


FIG. 5

13/49



**FIG. 6**  
**SUBSTITUTE SHEET (RULE 26)**

14/49

	L fragment	Cons-p9	L1* (A)	L2(A)	L6(A)	L8(A)	L9(A)	L3(B)	L7* (B)	L10(B)	L4(C)	Cons-D	Cons-F
Core	16	N	S	.	.	.	.	.	.	.	.	.	N
	36	L	.	.	.	P	.	.	.	.	.	.	L
	52	A	.	.	.	.	.	T	T	T	T	T	AT
	70	R	.	.	.	.	.	Q	Q	Q	.	R,Q	R,Q
	189	A	.	.	.	.	.	.	T	.	.	.	A
	195	R	.	.	.	.	.	H	.	H	.	.	R
	231	R	.	Q	Q	Q	.	.	.	.	.	.	R
E1	233	G	.	.	.	.	.	A	A	A	.	.	G
	234	N	.	.	.	.	.	D	D	D	.	.	N
	250	N	.	.	.	.	.	.	.	.	D	.	N
	299	E	A	.	.	.	A	.	.	.	.	.	E
	304	C	.	V	.	.	.	.	.	.	.	.	C
	379	A	.	.	.	.	.	T	.	T	.	.	A

FIG. 7A

15/49

	L fragment	Cons-p9	L1* (A)	L2(A)	L6(A)	L8(A)	L9(A)	L3(B)	L7* (B)	L10(B)	L4(C)	Cons-D	Cons-F
E2	384	E	.	.	.	.	.	T	T	T	.	E,I	E,I
	386	H	.	.	.	.	.	V	V	V	.	H,Y	H,Y
	388	T	.	.	.	.	.	S	S	S	.	I,S	I,S
	390	R	.	.	.	.	.	G	G	G	.	G	R,G
	391	V	.	.	.	.	.	.	.	A	.	.	V
	392	A	V	.	.	V	V	.	.	.	V	V	AV
	394	H	.	.	.	.	.	R	R	R	R	.	H
	405	S	.	.	.	.	.	.	P	.	.	.	S
	434	Q	.	.	.	.	.	H	H	H	.	H	Q,H
	438	F	.	.	.	.	.	L	L	L	L	L	F,L
	444	A	.	.	.	.	.	T	T	T	T	T	A,T
	450	S	.	.	.	.	.	.	.	.	P	.	S
	458	S	.	.	.	N	.	.	.	.	.	.	S
	466	A	.	.	.	.	.	V	V	V	.	AV	AV
	474	Y	.	.	.	.	.	H	.	.	.	.	Y
	476	K	.	.	.	.	.	E	E	E	E	E	K,E
	496	V	.	.	.	.	.	I	I	I	I	I	V,I
	524	V	.	.	.	.	A	.	A	.	.	.	V
	536	V	.	M	.	.	.	.	.	.	.	.	V
	580	I	.	.	.	.	.	.	V	.	.	.	I
	622	L	V	.	.	.	V	.	.	.	.	.	L
	673	Q	.	.	.	P	.	.	.	.	.	.	Q
	783	A	.	.	.	.	.	.	V	.	.	.	A
p7													

FIG. 7B

16/49

	L fragment	Cons-p9	L1*(A)	L2(A)	L6(A)	L8(A)	L9(A)	L3(B)	L7*(B)	L10(B)	L4(C)	Cons-D	Cons-F
NS2	820	G	.	.	.	.	.	.	S	.	.	.	G
	857	M	.	.	.	.	.	I	.	.	.	.	M
	927	K	.	.	.	.	.	.	R	.	.	.	K
	934	V	I	I	.	I	I	.	.	.	.	.	V
	937	A	.	.	V	.	.	.	.	.	.	.	A
	978	A	.	.	.	.	.	D	D	D	.	D	AD
	1028	P	.	.	.	S	.	.	.	.	.	.	P
	1031	A	.	.	.	.	.	.	.	T	.	.	A
	1043	V	.	.	I	.	.	I	I	I	.	I	VI
	1067	Q	.	.	.	.	.	H	H	H	.	Q,H	Q,H
NS3	1097	I	X	.	.	.	.	.	.	.	.	.	I
	1188	G	R	.	.	.	.	.	.	.	.	.	G
	1215	S	.	.	T	.	.	.	.	.	.	.	S
	1223	F	.	S	.	.	.	.	.	.	.	.	F
	1226	A	.	.	.	.	.	.	.	V	.	.	A
	1339	A	V	.	.	.	.	.	.	.	.	.	A
	1399	K	N	.	.	.	.	.	.	.	.	.	K
	1503	T	.	.	.	.	.	S	.	S	.	.	T
	1528	Y	.	.	.	.	.	.	.	.	H	.	Y
	1535	T	A	.	.	.	.	.	.	.	.	.	T
NS4A	1662	L	.	P	.	.	.	.	.	.	.	.	L

FIG. 7C

17/49

	L fragment	Cons-p9	L1*(A)	L2(A)	L6(A)	L8(A)	L9(A)	L3(B)	L7*(B)	L10(B)	L4(C)	Cons-D	Cons-F
NS4B	1753	K	.	R	.	.	.	.	.	.	.	.	K
	1805	H	.	.	N	.	.	N	.	N	N	N	H,N
	1949	S	.	.	.	.	.	.	.	.	P	.	S
	2105	M	.	.	.	.	V	I	.	I	.	.	M
	2136	K	.	.	.	.	.	.	.	.	R	.	K
NS5A	2146	T	.	.	.	.	.	A	A	A	.	IA	IA
	2226	L	.	.	.	.	.	P	.	.	.	.	L
	2259	L	.	.	.	.	.	F	.	.	.	.	L
	2262	E	.	.	.	.	.	D	D	D	.	ED	ED
	2334	V	.	.	.	.	.	I	.	.	.	.	V
	2371	L	.	.	.	.	.	Q	Q	Q	.	LQ	LQ
	2385	Y	.	.	.	.	.	.	.	.	H	.	Y
	2692	N	.	.	.	.	.	.	S	.	.	.	N
	2757	A	.	.	.	.	.	.	.	.	.	.	A
	2785	C	.	R	.	.	.	.	.	.	.	.	C
NS5B	2824	I	.	V	.	.	.	.	.	.	.	.	I
	2861	A	.	.	.	.	.	V	.	.	.	.	A
	S fragment		S5	S9	S2	S3	S7	S8	S10	S4	S6		
	2968	G	.	.	.	.	.	S	S	.	.	.	G
	2975	S	.	.	.	.	.	G	G	G	G	.	S
	2978	D	.	.	.	.	.	.	.	.	G	.	D
	2999	S	.	F	F	F	.	.	.	.	.	.	S
			.	.	.	.	.	.	.	.	.	.	.
			.	.	.	.	.	.	.	.	.	.	.
			.	.	.	.	.	.	.	.	.	.	.

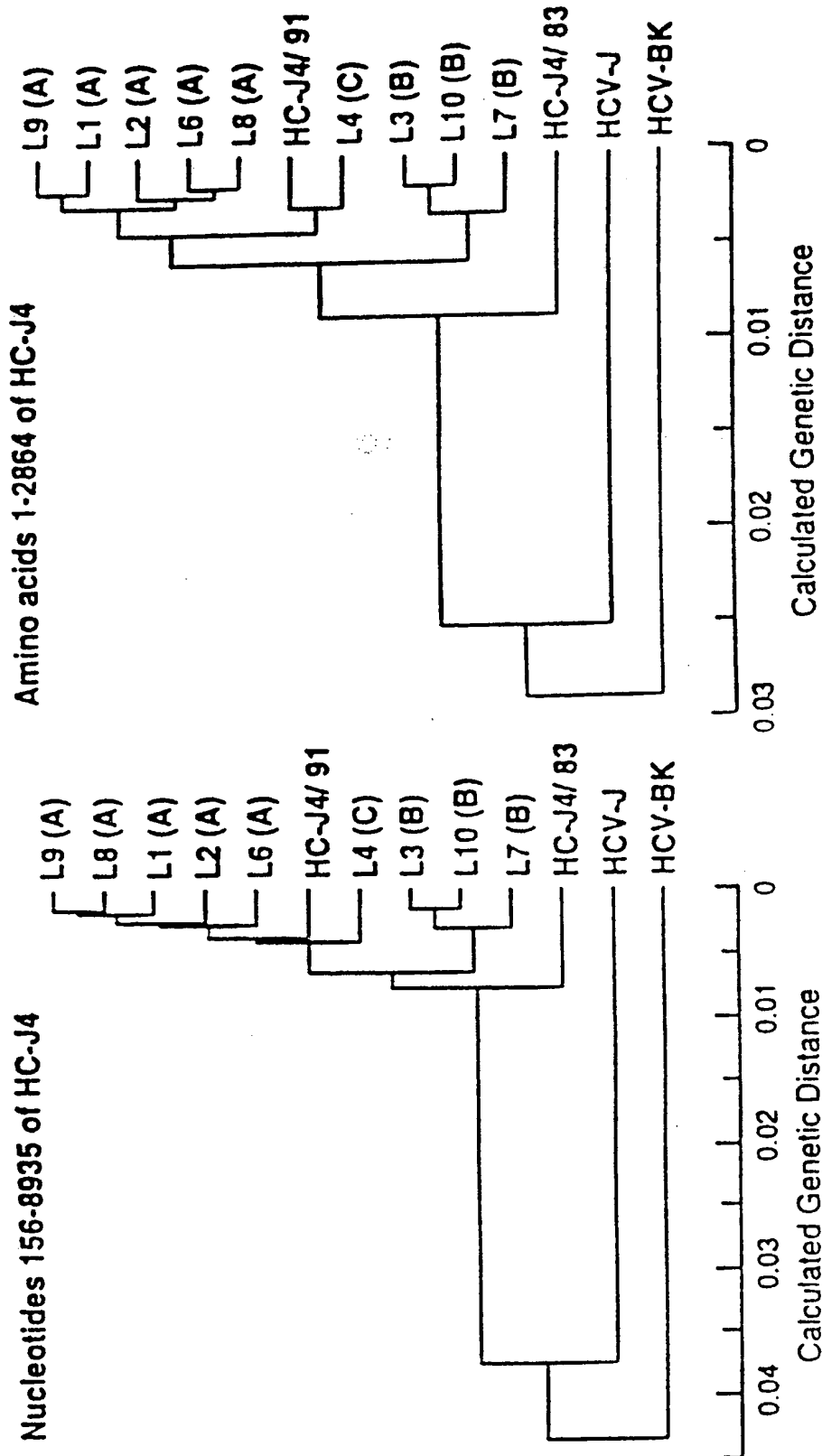
FIG. 7D

18/49

nt	L1 (A)	L2 (A)	L6 (A)	L8 (A)	L9 (A)	L3 (B)	L7 (B)	L10 (B)	L4 (C)	HC-J4/91	HC-J4/83
aa											
L1 (A)		0.56	0.60	0.36	0.33	1.50	1.53	1.46	0.95	0.83	1.79
L2 (A)	0.59		0.55	0.35	0.50	1.49	1.51	1.45	0.98	0.82	1.77
L6 (A)	0.52	0.42		0.31	0.55	1.33	1.38	1.29	0.80	0.68	1.58
L8 (A)	0.42	0.38	0.31		0.31	1.32	1.34	1.28	0.79	0.65	1.62
L9 (A)	0.35	0.52	0.45	0.35		1.42	1.42	1.38	0.91	0.75	1.66
L3 (B)	1.47	1.43	1.15	1.33	1.36		0.61	0.30	1.43	0.90	1.51
L7 (B)	1.36	1.33	1.05	1.22	1.22	0.66		0.57	1.47	0.95	1.54
L10 (B)	1.36	1.33	0.59	1.22	1.26	0.31	0.56		1.37	0.85	1.42
L4 (C)	0.77	0.80	0.59	0.63	1.26	1.12	1.08	1.01		0.76	1.73
HC-J4/91	0.94	0.91	0.63	0.80	0.87	0.77	0.73	0.66	0.52		1.22
HC-J4/83	1.96	1.89	1.68	1.85	1.82	1.75	1.61	1.61	1.71	1.40	

FIG. 8

19/49



**FIG. 9**



20/49

		379	413	468	486
HC-J4L6 (A)	:	AGVDG	ETHHTGRVAGHTTSGFTSLFSSGAS QKIQL	GWGPIT	YTKPNSS DQRPYC
HC-J4L2 (A)	:				
HC-J4/91-20	:		R.		E.
HC-J4L1 (A)	:		V.		
HC-J4L8 (A)	:		V.		
HC-J4L9 (A)	:		V.		G.
HC-J4/91-21	:		V.		
HC-J4L4 (C)	:		V.R.		E.
HC-J4/91-23	:		V.R.		E.
HC-J4/91-22	:		V.R.	A.	E.
HC-J4L7 (B)	:		T.Y.S.G. R. P.		E.
HC-J4L10(B)	:	T.	T.Y.S.GA. R.		E.
HC-J4L3 (B)	:	T.	T.Y.S.G. R.	H.E.	
HC-J4/91-26	:	T.	T.Y.S.G. R.		G.D.L.
HC-J4/91-25	:		A.Y.S.G. R.		E.
HC-J4/91-24	:		A.Y.S.G. R.		E..P
HC-J4/91	:		A.Y.S.G. R.		E..P
HC-J4/91-27	:		K.Y.S.GA.S. R. P. R.		ESG.R
HC-J4/83	:		Y.S.GA.S. TLA. P. R.		E.D.P


  
 HVR2



  
 HVR1

FIG. 10

21/49

## 3' Untranslated Region

1  
 IC-J4 :GCCAGCCCCC GATTGGGGG GACACTCCAC CATAGATCAC TCCCCTGTGA GGAAGTACTG TCTTCACGCA GAAAGCGTCT AGCCATGGCG 90  
 CV-J4L6S :TGA..... GA.....  
 CV-H77C :TGA..... GA.....  
 180  
 91  
 IC-J4 :TTAGTATGAG TGTCGTGCAG CCTCCAGGAC CCCCCCTCCC GGGAGAGCCA TAGTGGTCTG CGGAACCGGT GAGTACACCG GAATTGCCAG  
 CV-J4L6S :.....  
 CV-H77C :..... Pin Al  
 270  
 181  
 HC-J4 :GACGACCGGG TCCTTCTTGG GATCAACCCG CTCAATGCCT GGAGATTGG GCGTGCCCCC GCGAGACTGC TAGCCGAGTA GTGTGGGTC  
 pCV-J4L6S :..... A.....  
 pCV-H77C :.....  
 341  
 271  
 HC-J4 :GCGAAGGCC TTGTGGTACT GCCTGATAGG GTGCTTGGGA GTGCCCGGG AGGTCTCGTA GACCGTGAC C  
 pCV-J4L6S :.....  
 pCV-H77C :.....

## 3' Untranslated Region

9372  
 HC-J4 :TGAACGGGA GCTAACCACT CCAGGCCAAT AGGCCTT--C CTG poly (U-UC)<sub>n</sub> 9513  
 pCV-J4L6S :..... -T..A...A.TT. ... poly (U-UC)<sub>81</sub>  
 pCV-H77C :..... G.C.TCT..A...A.TT. ... poly (U-UC)<sub>81</sub>  
 Bfr 1

3' variable region      poly U-UC region      3' variable region

## 3' conserved region (Cont.)

9514  
 H77 :CCCTAGTCAC GGCTAGCTGT GAAAGGTCCG TGAGCCGCAT GACTGCAGAG AGTGCTGATA CTGGCCCTCTC TGCAGATCAT GT 9595  
 pCV-J4L6S :.....  
 pCV-H77C :.....

FIG.11

22/49

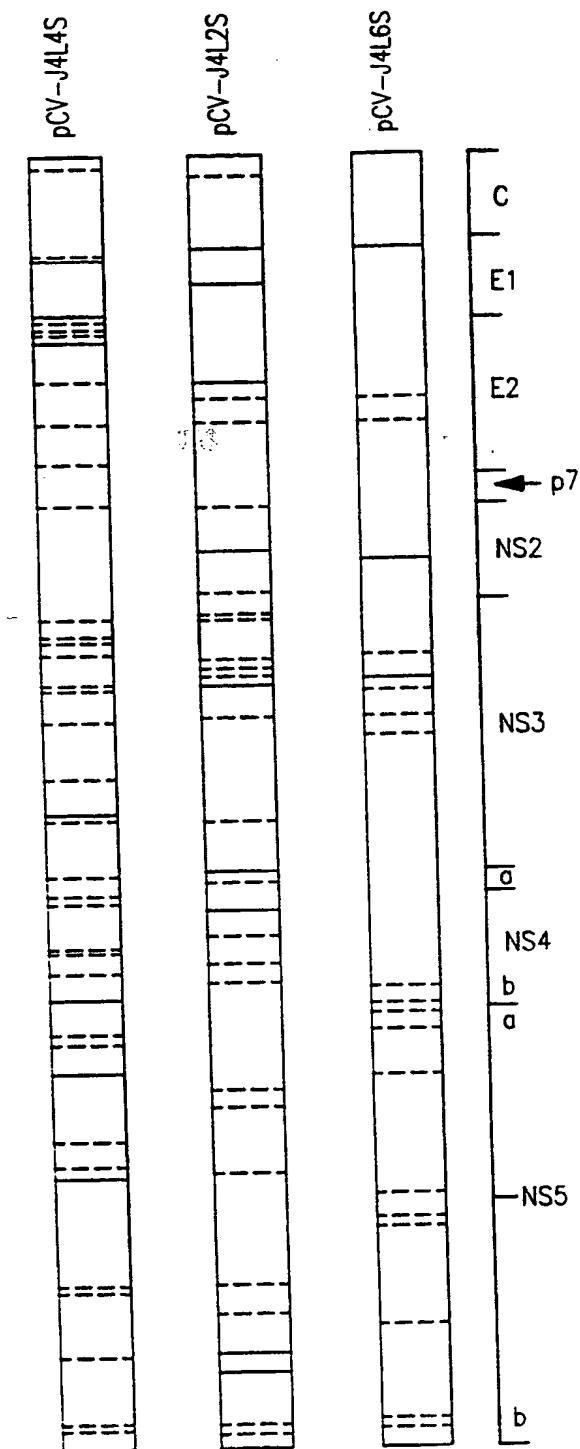


FIG. 12

SUBSTITUTE SHEET (RULE 26)

23/49

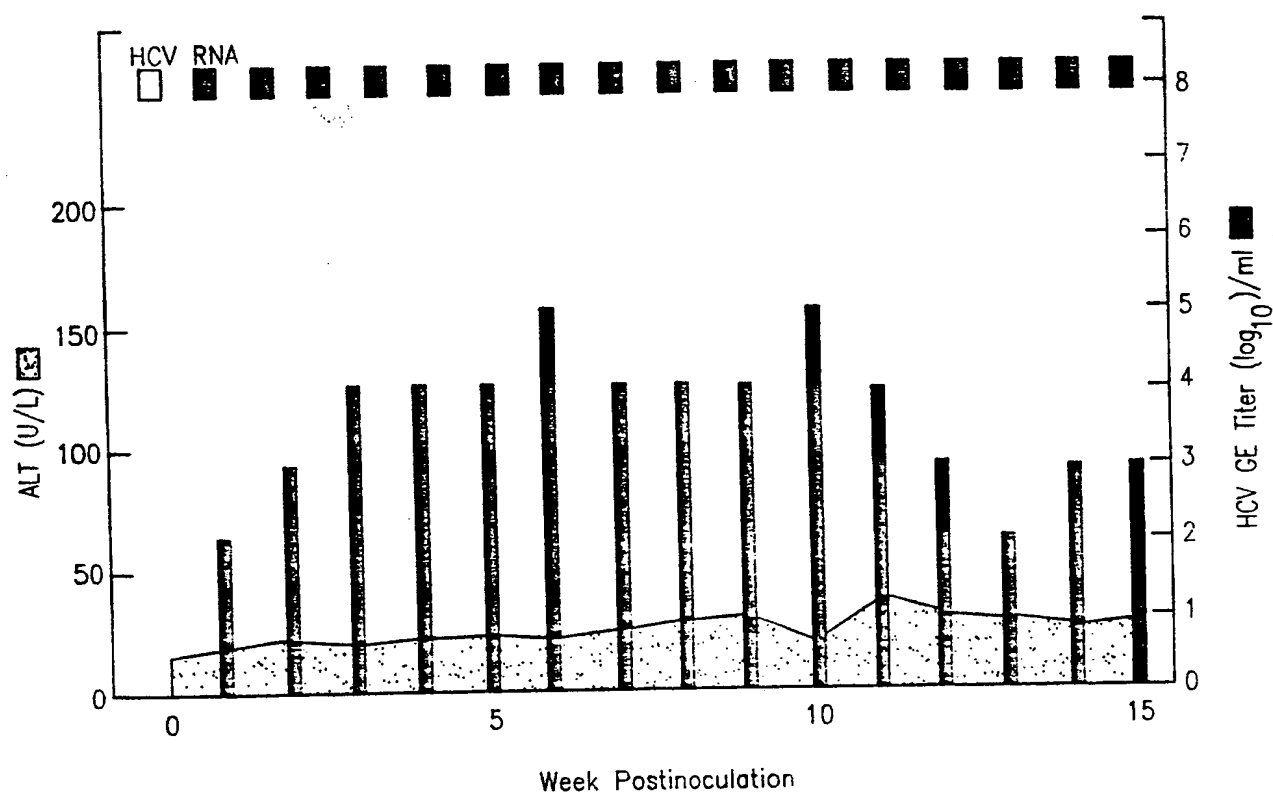


FIG. 13

24/49

## HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCCAGCCCCC	TGATGGGGGC	GACACTCCAC	CATGAATCAC	TCCCCGTGTA	50
GGAAC TACTG	TCTTCAACGA	GAAAGCGTCT	AGCCATGGCG	TTAGTATGAG	100
TGTCTGTGAG	CCTCCAGGAC	CCCCCCTCCC	GGGAGAGCCA	TAGTGGTCTG	150
CGGAACCGGT	GAGTACACCG	GAATTGOCAG	GACGACCGGG	TCCTTTCTTG	200
GATCAACCCG	CTCAATGCCT	GGAGATTTGG	GCGTGCCCCC	GCGAGACTGC	250
TAGCCGAGTA	GTGTTGGGTC	GCGAAAGGCC	TTGTGGTACT	GCCTGATAGG	300
GTGCTTGCGA	GTGCCCCGGG	AGGTCCTGTA	GACCGTGCAC	CATGAGCACG	350
AATCCTAAAC	CTCAAAGAAA	AACCAAAAGT	AACACCAACC	GCCGCCACAA	400
GGACGTCAAG	TTCCCGGGCG	GTGGTCAGAT	CGTTGGTGGG	GTTTACCTGT	450
TGCCCGCCAG	GGGCCCCAGG	TTGGGTGTGC	GCGCGACTAG	GAAGGCTTCC	500
GAGCGGTGCG	AACCTCGTGG	AAGGCGACAA	CCTATCCCAA	AGGCTCGCCG	550
ACCCGAGGGC	AGGGCCTGGG	CTCAGCCCGG	GTACCCCTTG	CCCCCTATG	600
GCAATGAGGG	CCTGGGGTGG	GCAGGATGGC	TCCTGTACAC	CCGCGGCTCC	650
CGGCCCTAGTT	GGGGCCCCAC	GGACCCCCCG	CGTAGGTGCG	GTAACCTGGG	700
TAAGGTCATC	GATACCTTAA	CATGCGGCTT	CGCCGATCTC	ATGGGGTACA	750
TTCCGCTCGT	CGGCGCCCCC	CTAGGGGGCG	CTGCCAGGGC	CTTGGCACAC	800
GGTGTCCGGG	TTCTGGAGGA	CGGCGTGAAC	TATGCAACAG	GGAACCTGCC	850
CGGTGTCTCT	TTCTCTATCT	TCCTCTTGGC	TCTGTCTGTC	TGTTTGACCA	900
TCCCAGCTTC	CGCTTATGAA	GTGCGCAACG	TGTCCGGGAT	ATACCATGTC	950
ACGAACGACT	GCTCCAACTC	AAGCATTGTG	TATGAGGCAG	CGGACGTGAT	1000
CATGCATACT	CCCGGGTGGG	TGCCCTGTGT	TCAGGAGGGT	AACAGCTCCC	1050
GTTGCTGGGT	AGCGCTCACT	CCCACGCTCG	CGGCCAGGAA	TGCCAGCGTC	1100
CCCACTACGA	CAATACGACG	CCACGTCGAC	TTGCTCGTTG	GGACGGCTGC	1150
TTTCTGCTCC	GCTATGTACG	TGGGGGATCT	CTGCCGATCT	ATTTTCTCTG	1200
TCTCCCAGCT	GTTACCTTTC	TGCCTCGCC	GGCATGAGAC	AGTGCAGGAC	1250
TGCAACTGCT	CAATCTATCC	CGGCCATGTA	TCAGGTACAC	GCATGGCTTG	1300
GGATATGATG	ATGAACTGGT	CACCTACAAC	AGCCCTAGTG	GTGTGCGAGT	1350
TGCTCCGGAT	CCCACAAGCT	GTGTTGGACA	TGGTGGCGGG	GGCCCACTGG	1400
GGAGTCCCTG	CGGGCCTTGC	CTACTATTCC	ATGGTAGGGA	ACTGGGCTAA	1450
GGTTCTGATT	GTGGCGCTAC	TCCTTGGCCG	CGTTGACGGG	GAGACCCACA	1500
CGACGGGGAG	GGTGGCCGGC	CACACCACCT	CCGGGTTTAC	GTCCCTTTTC	1550
TCATCTGGGG	CGTCTCAGAA	AATCCAGCTT	GTGAATACCA	ACGGCAGCTG	1600
GCACATCAAC	AGGACTGCCC	TAAATTGCAA	TGACTCCCTC	CAAACTGGGT	1650
TCTTTGCCGC	GCTGTTTTAC	GCACACAAGT	TCAACTCGTC	CGGGTGCCCG	1700
GAGCGCATGG	CCAGCTGCCG	CCCCATTGAC	TGGTTCCGCC	AGGGGTGGGG	1750
CCCCATCACC	TATACTAAGC	CTAACAGCTC	GGATCAGAGG	CCTTATTGCT	1800
GGCATTACGC	GCCTCGACCG	TGTGGTGTGC	TACCCGCGTC	GCAGGTGTGT	1850
GGTCCAGTGT	ATTGTTTTCAC	CCCAAGCCCT	GTGTGTGGTG	GGACCACCGA	1900

FIG. 14A

SUBSTITUTE SHEET (RULE 26)

## HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCGTTCCGGT	GTCCCTACGT	ATAGCTGGGG	GGAGAATGAG	ACAGACGTGA	1950
TGCTCCTCAA	CAACACGGGT	CCGCCACAAG	GCAACTGGTT	CGGCTGTACA	2000
TGGATGAATA	GTACTGGGTT	CACTAAGACG	TGCGGAGGTC	CCCGTGTAA	2050
CATCGGGGGG	GTOGGTAACC	GCACCTTGAT	CTGCCCCAAG	GACTGCTTCC	2100
GGAAGCACCC	CGAGGCTACT	TACACAAAT	GTGGCTCGGG	GOCCTGGTTG	2150
ACACCTAGGT	GCCTAGTAGA	CTACCCATAC	AGGCTTTGGC	ACTACCCCTG	2200
CACCTCTCAAT	TTTTCATCT	TTAAGGTTAG	GATGTATGTG	GGGGCGGTGG	2250
AGCACAGGCT	CAATGCCGCA	TGCAATTGGA	CTCGAGGAGA	GCGCTGTAAAC	2300
TTGGAGGACA	GGGATAGGTC	AGAACTCAGC	CCGCTGCTGC	TGTCTACAAC	2350
AGAGTGGCAG	ATACTGCCCT	GTGCTTTTAC	CACCCCTACG	GCTTTTATCCA	2400
CTGGTTTGAT	CCATCTCCAT	CAGAACATCG	TGGACGTGCA	ATACCTGTAC	2450
GGTGTAGGGT	CAGCGTTTGT	CTCCTTTTGA	ATCAAATGGG	AGTACATCCT	2500
GTGCTTTTTC	CTTCTCCTGG	CAGACGCGCG	CGTGTGTGCC	TGCTTGTGGA	2550
TGATGCTGCT	GATAGCCAG	GCTGAGCCCG	CCTTAGAGAA	CTTGGTGGTC	2600
CTCAATGCCG	CGTCCGTGGC	CGGAGCGCAT	GGTATTTCTCT	CCTTTCTTGT	2650
GTCTTTCTGC	GCCGCCCTGGT	ACATTAAAGG	CAGGCTGGCT	CCTGGGCGGG	2700
CGTATGCTTT	TTATGGCGTA	TGGCCGCTGC	TCCTGCTCCT	ACTGGCGTTA	2750
CCACCACGAG	CTTACGCCTT	GGACCGGGAG	ATGGCTGCAT	CGTGGCGGGG	2800
TGCGGTTCTT	GTAGGTCTGG	TATTCTTGAC	CTTGTACCA	TACTACAAAG	2850
TGTTTCTCAC	TAGGCTCATA	TGGTGGTTAC	AATACTTTAT	CACCAGAGCC	2900
GAGCGGCACA	TGCAAGTGTG	GGTCCCCCCC	CTCAACGTTT	GGGGAGGCGG	2950
CGATGCCATC	ATCCTCCTCA	CGTGTGCGGT	TCATCCAGAG	TTAATTTTGT	3000
ACATCACCAA	ACTCCTGCTC	GCCATACTCG	GCCCGCTCAT	GGTGCTCCAG	3050
GCTGGCATAA	CGAGAGTGCC	GTACTTCTGT	CGCGCTCAAG	GGCTCATTCG	3100
TGCATGCATG	TTAGTGGGAA	AAGTCGCGCG	GGGTCAATTAT	GTCCAAATGG	3150
TCTTCATGAA	GCTGGGCGCG	CTGACAGGTA	CGTACGTTTA	TAACCATCTT	3200
ACCCCACTGC	GGGACTGGGC	CCACGCGGGC	CTACGAGACC	TTGCGGTGGC	3250
GGTAGAGCCC	GTCGTCTTCT	CCGCCATGGA	GACCAAGGTC	ATCACCTGGG	3300
GAGCAGACAC	CGCTGCGTGT	GGGGACATCA	TCTTGGGTCT	ACCGTCTCC	3350
GCCCCAAGGG	GGAAGGAGAT	ATTTTGTGGA	CCGCTGATA	GTCTCGAAGG	3400
GCAAGGGTGG	CGACTCCTTG	CGCCCATCAC	GGCCTACTCC	CAACAAACGC	3450
GGGGCGTACT	TGGTTGCATC	ATCACTAGCC	TCACAGGCCG	GGACAAGAAC	3500
CAGGTCTGAAG	GGGAGGTTCA	AGTGGTTTCT	ACCGCAACAC	AATCTTTCTT	3550
GGCGACCTGC	ATCAACGGCG	TGTGCTGGAC	TGTCTACCAT	GGCGCTGGCT	3600
CGAAGACCTT	AGCCGGTCCA	AAAGGTCCAA	TCACCCAAAT	GTACACCAAT	3650
GTAGACCTGG	ACCTCGTCCG	CTGGCAGGCG	CCCCCGGGG	CGCGCTCCAT	3700
GACACCATGC	AGCTGTGGCA	GCTCGGACCT	TTACTTGGTC	ACGAGACATG	3750
CTGATGTICAT	TCCGGTGGCG	CGGCGAGGCG	ACAGCAGGGG	AAGTCTACTC	3800

FIG. 14B

SUBSTITUTE SHEET (RULE 26)

## HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCCCCAGGC	CCGTCCTCTA	CCTGAAAGGC	TOCTCGGGTG	GTCCATTCGT	3850
TTGCCCCCTG	GGGCACGTG	TGGGGGCTTT	COGGGCTGCT	GIGTGCACCC	3900
GGGGGGTGG	GAAGGGGGTG	GACTTCATAC	COGTTGAGTC	TATGGAAACT	3950
ACCATGGGGT	CTCCGGTCTT	CACAGACAAC	TCAACCCCCC	CGGCTGTACC	4000
GCAGACATT	CAAGTGGCAC	ATCTGCACGC	TOCTACTGGC	AGGGGCAAGA	4050
GCACCAAGT	GCCGGCTGG	TATGCAGCCC	AAGGGTACAA	GGTGCTGGTC	4100
CTGAACCCGT	CCGTTGGCGC	CACTTACGGG	TTTGGGGGGT	ATATGTCCAA	4150
GGCACACGGT	ATCGACCCCTA	ACATCAGAAC	TGGGGTAAAG	ACCATTACCA	4200
CGGGGGGCTC	CATTACGTAC	TCCACCTATG	GCAAGTTCTT	TGCGGACGGT	4250
GGCTGTCTTG	GGGGGGCTTA	TGACATCATA	ATATGTGATG	AGTGCCACTC	4300
AACTGACTCG	ACTACCATCT	TGGGCATGG	CACAGTCTTG	GACCAAGGG	4350
AGACGGCTGG	AGCGGGGCTC	GTGGTGCTCG	CCACCGCTAC	ACCTCGGGA	4400
TCCGTTACCG	TGCCACACCC	CAATATCGAG	GAAATAGGOC	TGTCCAACAA	4450
TGGAGAGATC	CCCTTCTATG	GCAAAGCCAT	CCCATTTGAG	GCCATCAAGG	4500
GGGGGAGGCA	TCTCATTTTC	TGCCATTCCA	AGAAGAAATG	TGACGAGCTC	4550
GCCGCAAAGC	TGACAGGCGT	CGGACTGAAC	GCTGTAGCAT	ATTACCGGG	4600
CCTTGATGTG	TCCGTCATAC	CGCCTATCGG	AGACGTGCTT	GTGGTGCCAA	4650
CAGACGCTCT	AATGACGGGT	TTCACCGGGG	ATTTTGACTC	AGTGATCGAC	4700
TGCAATACAT	GTGTACCCCA	GACAGTCGAC	TTCAGCTTGG	ATCCACCTT	4750
CACCATTTGAG	ACGACGACCG	TGCCCCAAGA	CGCGGTGTGG	CGCTCGCAAC	4800
GGCGAGGTAG	AACTGGCAGG	GGTAGGAGTG	GCATCTACAG	GTTTGTGACT	4850
CCAGGAGAAC	GGCCCTCGGG	CATGTTGGAT	TCTTCGGTCC	TGTTGTGAGTG	4900
CTATGACGGG	GGCTGTGCTT	GGTATGAGCT	CACGCCCCGT	GAGACCTCGG	4950
TTAGGTTGGG	GGCTTACCTA	AATACACCAG	GGTTGCCCCG	CTGCCAGGAC	5000
CATCTGGAGT	TCTGGGAGAG	CGTCTTCACA	GGCCTCACCC	ACATAGATGC	5050
CCACTTCCCTG	TCCCAGACTA	AACAGGCAGG	AGACAACTTT	CCTTACCTGG	5100
TGGCATATCA	AGCTACAGTG	TGCGCCAGGG	CTCAAGCTCC	ACCTCCATCG	5150
TGGGACCAAA	TGTGGAAGTG	TCTCATACGG	CTGAAACCTA	CACCTGCACGG	5200
GCCAACACCC	CTGCTGTATA	GGCTAGGAGC	CGTCCAAAAT	GAGGTCAATCC	5250
TCACACACCC	CATAACTAAA	TACATCATGG	CATGCATGTC	GGCTGAOCTG	5300
GAGGTGCTCA	CTAGCACCTG	GGTGCTGGTA	GGCGGAGTCC	TTGCAGCTTT	5350
GGCCGCATAC	TGCCTGACGA	CAGGCAGTGT	GGTCATTGTG	GGCAGGATCA	5400
TCTTGTCCGG	GAAGCCAGCT	GTGGTTCCCG	ACAGGGAAGT	CCTCTACCAG	5450
GAGTTGATG	AGATGGAAGA	GTGTGCTCA	CAACTTCCTT	ACATCGAGCA	5500
GGGAATGCAG	CTCGCCGAGC	AATTCAAGCA	AAAGGCGCTC	GGGTTGTGTC	5550
AAACGGCCAC	CAAGCAAGCG	GAGGCTGCTG	CTCCCGTGGT	GGAGTCCAAG	5600
TGCGGAGCCC	TTGAGACCTT	CTGGGCGAAG	CACATGTGGA	ATTTTCATCAG	5650
CGGAATACAG	TACCTAGCAG	GCTTATCCAC	TCTGCCTGGA	AACCCCGCGA	5700

FIG. 14C

SUBSTITUTE SHEET (RULE 26)

27/49

## HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TAGCATCATT	GATGGCATT	ACAGCTTCTA	TCACTAGCCC	GCTCACCACC	5750
CAAAACACCC	TCCGTGTTAA	CATCTTGGGG	GGATGGGTTG	CTGCCCAACT	5800
CGCTCCTCCC	AGCGCTGCGT	CAGCTTTTGG	GGCGCGCGGC	ATCGCGCGAG	5850
CGGCTGTGTT	CAGCATAGGC	CTTGGGAAGG	TGCTCGTGGG	CATCTTGGCG	5900
GGCTATGGGG	CAGGGGTAGC	CGGCGCACTC	GTTGGCTTTA	AGGTTCATGAG	5950
CGGCGAGGTG	CCCTCCACCG	AGGACCTGGT	CAACTTACTC	CCTGCCATCC	6000
TCTCTCCTGG	TGCGCTGGTC	GTCGGGGTGG	TGTGCGCAGC	AATACTGGGT	6050
CGGCACGTGG	GCGCGGGAGA	GGGGGCTGTG	CAGTGGATGA	ACCGGCTGAT	6100
AGCGTTGGCT	TGCGGGGGTA	AACAGTCTTC	CCCTACGCAC	TATGTGGCTG	6150
AGAGCGACGC	TGCAGCACGT	GTCACTCAGA	TCTCTCTTAG	CCTTACCATC	6200
ACTCAACTGC	TGAAGCGGCT	CCACCAGTGG	ATTAATGAGG	ACTGCTCTAC	6250
GCCATGCTCC	GGCTCGTGGC	TAAGGGATGT	TTGGGATTGG	ATATGCACGG	6300
TGTTGACTGA	CTTCAAGACC	TGGCTCCAGT	CCAAACTCCT	GCGCGGGTTA	6350
CCGGGAGTCC	CTTTCCGTGC	ATGCCAACGC	GGGTACAAGG	GAGTCTGGCG	6400
GGGGGACGGC	ATCATGCAAA	CCACCTGCCC	ATGCGGAGCA	CAGATCGCCG	6450
GACATGTCAA	AAACGGTTCC	ATGAGGATCG	TAGGGCCTAG	AACCTGCAGC	6500
AACACGTGGC	ACGGAACGTT	CCCCATCAAC	GCATACACCA	CGGACCTTGG	6550
CACACCTCC	CCGGCGCCCA	ACTATTCCAG	GGCGCTATGG	CGGGTGGCTG	6600
CTGAGGAGTA	CGTGGAGGTT	ACGCGTGTGG	GGGATTTCCA	CTACGTGACG	6650
GGCATGACCA	CTGACAACGT	AAAGTGCCCA	TGCCAGGTTT	CGGCCCCCGA	6700
ATTCTTACAG	GAGGTGGATG	GAGTGGCGGT	GCACAGGTAC	GCTCGGGCGT	6750
GCAAACCTCT	TCTACGGGAG	GACGTCACGT	TCCAGGTCCG	GCTCAACCAA	6800
TACTTGGTCC	GGTCGCAGCT	CCCATGCGAG	CCCGAACCGG	ACGTAACAGT	6850
GCTTACTTCC	ATGCTCACCG	ATCCCTCCCA	CATTACAGCA	GAGACGGCTA	6900
AGCGTAGGCT	GGCTAGAGGG	TCTCCCCCCT	CTTTAGCCAG	CTCATCAGCT	6950
AGCCAGTTGT	CTGCGCCTTC	TTTGAAGGCG	ACATGCACTA	CCCACCATGA	7000
CTCCCCGGAC	GCTGACCTCA	TGAGGGCCAA	CCTCTTGTGG	CGGCAGGAGA	7050
TGGGCGGAAA	CATCACTCGC	GTTGAGTCAG	AGAATAAGGT	AGTAATTCTG	7100
GACTCTTTTC	AACCGCTTCA	CGCGGAGGGG	GATGAGAGGG	AGATATCCGT	7150
CGCGGCGGAG	ATCCTGCGAA	AATCCAGGAA	GTTCCCTTCA	GCGTTGCCCA	7200
TATGGGCACG	CCCGGACTAC	AATCCTCCAC	TGCTAGAGTC	CTGGAAGGAC	7250
CCGGACTACG	TCCCTCCGGT	GGTACACGGA	TGCCCATTGC	CACCTACCAA	7300
GGCTCCTCCA	ATACCACCTC	CACGGAGAAA	GAGGACGGTT	GTCCTGACAG	7350
AATCCAATGT	GTCCTCTGCC	TTGGCGGAGC	TGCGCACTAA	GACCTTCCGT	7400
AGCTCCGGAT	CGTCCGGCGT	TGATAGCGGC	ACGGCGACCG	CCCTTCTCTG	7450
CCTGGCCTCC	GACGACGGTG	ACAAAGGATC	CGACGTTGAG	TGCTACTCCT	7500
CCATGCCCCC	CCTTGAAGGG	GAGCCGGGGG	ACCCCGATCT	CAGCGACGGG	7550
TCTTGGTCTA	CCGTGAGTGA	GGAGGCTAGT	GAGGATGTGG	TCTGCTGCTC	7600

FIG. 14D

SUBSTITUTE SHEET (RULE 26)



28/49

HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AATGTCTTAT	ACGTGGACAG	GCGCCCTGAT	CAAGCCATGC	GCTGCGGAGG	7650
AAAGTAAGCT	GCCCATCAAC	CGGTGAGCA	ACTCTTTGCT	GCGTCACCAC	7700
AACATGGTCT	ACGCCACAAC	ATCCCGCAGC	GCAAGCCTCC	GGCAGAAGAA	7750
GGTCACCTTT	GACAGATTGC	AAGTCTTGGA	TGATCATTAC	CGGGACGTAC	7800
TCAAGGAGAT	GAAGGCGAAG	GCGTCCACAG	TTAAGGCTAA	GCTTCTATCT	7850
ATAGAGGAGG	CCTGCAAGCT	GACGCCCCCA	CATTGCGCCA	AATCCAAATT	7900
TGGCTATGGG	GCAAAGGACG	TCCGGAACCT	ATCCAGCAGG	GCCGTAAACC	7950
ACATCCGCTC	CGTGTGGGAG	GACTTGCTGG	AAGACACTGA	AACACCAATT	8000
GACACCACCA	TCATGGCAAA	AAGTGAGGTT	TTCTGCGTCC	AACCAGAGAA	8050
GGGAGGCCGC	AAGCCAGCTC	GCCTTATCGT	ATTCCCAGAC	CTGGGAGTTC	8100
GTGTATGCGA	GAAGATGGCC	CTTTACGACG	TGGTCTCCAC	CCTTCTCAG	8150
GCCGTGATGG	GCTCCTCATA	CGGATTTCAA	TACTCCCCCA	AGCAGCGGGT	8200
CGAGTTCCTG	GTGAATACCT	GGAAATCAAA	GAAATGOCCT	ATGGGCTTCT	8250
CATATGACAC	CCGCTGTTTT	GACTCAACGG	TCACTGAGAG	TGACATTGGT	8300
GTTGAGGAGT	CAATTTACCA	ATGTTGTGAC	TTGGCCCCCG	AGGCCAGACA	8350
GGCCATAAGG	TGGCTCACAG	AGCGGCTTTA	CATCGGGGGT	CCCCTGACTA	8400
ACTCAAAGG	GCAGAACTGC	GGTTATCGCC	GGTGCCGCGC	AAGTGGCGTG	8450
CTGACGACTA	GCTGCGGTAA	TACCTTCACA	TGTTACTTGA	AGGCCACTGC	8500
AGCCTGTGGA	GCTGCAAAGC	TCCAGGACTG	CACGATGCTC	GTGAACGGAG	8550
ACGACCTTGT	CGTTATCTGT	GAAAGCGCGG	GAACCCAGGA	GGATGCGGCG	8600
GCCCTACGAG	CCTTCACGGA	GGCTATGACT	AGGTATTCCG	CCCCCCCCCG	8650
GGATCCGCCC	CAACCAGAAT	ACGACCTGGA	GCTGATAACA	TCATGTTTCT	8700
CCAATGTGTC	AGTCGCGCAC	GATGCATCTG	GCAAAGGGT	ATACTACCTC	8750
ACCCGTGACC	CCACCACCCC	CCTTGCACGG	GCTGCGTGGG	AGACAGCTAG	8800
ACACACTCCA	ATCAACTCTT	GGCTAGGCCA	TATCATCATG	TATGCGCCCA	8850
CCCTATGGGC	AAGGATGATT	CTGATGACTC	ACTTTTCTTC	CATCCTTCTA	8900
GCTCAAGAGC	AACTTGAAAA	AGCCCTGGAT	TGTCAGATCT	ACGGGGCTTG	8950
CTACTCCATT	GAGCCACTTG	ACCTACCTCA	GATCATTGAA	CGACTCCATG	9000
GTCTTAGCGC	ATTTCACACTC	CACAGTACT	CTCCAGGTGA	GATCAATAGG	9050
GTGGCTTCAT	GCCTCAGGAA	ACTTGGGGTA	CCACCCCTGC	GAACCTGGAG	9100
ACATCGGGCC	AGAAGTGTCC	GCGCTAAGCT	ACTGTCCAG	GGGGGGAGGG	9150
CCGCCACTTG	TGGCAGATAC	CTCTTTAACT	GGGCAGTAAG	GACCAAGCTT	9200
AAACTCACTC	CAATCCCGGC	CGCGTCCAG	CTGGACTTGT	CTGGCTGGTT	9250
CGTCCCTGGT	TACAGCGGGG	GAGACATATA	TCACAGCCTG	TCTCGTGCOC	9300
GACCCCGCTG	GTTCGCGTGT	TGCCTACTCC	TACTTTCTGT	AGGGGTAGGC	9350
ATTTACCTGC	TCCCCAACCG	ATGAACGGGG	AGCTAACCAC	TCCAGGCCCTT	9400
AAGCCATTTC	CTGTTTTTTT	TTTTTTTTTT	TTTTTTTTTT	TCTTTTTTTT	9450
TTTCTTTCT	TTCCTTCTTT	TTTTCTTTTC	TTTTTCCCTT	CTTTAATGGT	9500

FIG. 14E

SUBSTITUTE SHEET (RULE 26)

29/49

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GGCTCCATCT	TAGCCCTAGT	CACGGCTAGC	TGTGAAAGGT	CCGTGAGCCG	9550
CATGACTGCA	GAGAGTGCTG	ATACTGGCCT	CTCTGCAGAT	CATGT	9595

FIG. 14F

30/49

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
MSINPKPQRK	TKRNINRRPQ	DKFPGGGQI	VGGVYLLPRR	GPRLGVRATR	50
KASERSQPRG	RRQPIPKARR	PEGRAWAQP	YPWPLYGNEG	LGWAGWLLSP	100
RGSRPSWGPT	DPRRRSRNLG	KVIDILTQGF	ADLMGYIPLV	GAPLGGAARA	150
LAHGVRLLED	GVNYATGNLP	GCSFSIFLLA	LLSCLTIPAS	AYEVRNWSGI	200
YHVINDCSNS	SIVYEAADVI	MHTPGCVPCV	QEGNSSRCW	ALTPTLAARN	250
ASVPTTITRR	HVDLLVGTA	FCSAMYVGL	CGSIFLVSQL	FTFSPPRHET	300
VQDCNCSTYP	GHVSGHRMAW	DMMNWSPTT	ALVVSQLLRI	PQAVVDMVAG	350
AHWGVLAGLA	YYSMVGNWAK	VLIVALLFAG	VDGETHTTGR	VAGHTTSGFT	400
SLFSSGASQK	IQLVNINGSW	HINRIALN	DSLQTGFFAA	LFYAHKFNSS	450
GCPERMASCR	PIDWFAQGW	PITYTKPNSS	DQRPYCWHYA	PRPGVVPAS	500
QVCGPVYCFT	PSPVAVGTID	RSGVPTYSWG	ENETDVMLIN	NIRPPQGNWF	550
GCTWMNSTGF	TKTCGGPPCN	IGGVGNRTLI	CPIDCFRKHP	EATYTKCGSG	600
PWLTPRCLVD	YPYRLWHYPC	TLNFSIFKVR	MYVGGVEHRL	NAACNWIRGE	650
RONLEDRRS	ELSPILLSTT	EWQILPCAFT	TLPALSTGLI	HLHQNIVDVQ	700
YLYGVGSAFV	SFAIKWEYIL	LLFLLLADAR	VCACIWMMLL	IAQAEAALEN	750
LVLINAASVA	GAHGILSFLV	FFCAAWYIKG	RLAPGAAYAF	YGVWPLLLLL	800
LALPPRAYAL	DREMAASCGG	AVLVGLVFLT	LSPYYKVFLT	RLIWWLQYFI	850
TRAEAHMQVW	VPPLNVRGGR	DAIILLTCAV	HPELIFDITK	LLAILGLPLM	900
VLQAGITRVP	YFVRAQGLIR	ACMLVRKVAG	GHYVQMVFMK	LGALTGTIVY	950
NHLTPLRDWA	HAGLRDLAVA	VEPVVFSAME	TKVTTWGADT	AACGDIILGL	1000
PVSARRGEI	FLGPADSLEG	QGWRLAPIT	AYSQQTRGVL	GCIITSLTGR	1050
DKNQVEGEVQ	VVSTATQSFL	ATCINGVCWT	VYHGAGSKTL	AGPKGPITQM	1100
YINVDLDELVG	WQAPPGARSM	TPCSCGSSDL	YLVTRHADVI	PVRRRGDSRG	1150
SLLSPRPVSY	LKGSSGGPLL	CPSGHVGVF	RAAVCTRGVA	KAVDFIPVES	1200
METIMRSPVF	TDNSTPPAVP	QTFQVAHLHA	PTGSGKSTKV	PAAYAAQGYK	1250
VLVLNPSVAA	TLGFGAYMSK	AHGIDPNIRT	GVRTTTTGGS	ITYSTYKFL	1300
ADGGCSGGAY	DIICDECHS	TDSTTILGIG	TVLDQAETAG	ARLVVLATAT	1350
PPGSVTVPHP	NIEEIGLSNN	GEIPFYGKAI	PIEATKGRH	LIFCHSKKKC	1400
DELAACKLTGL	GLNAVAYYRG	LDVSVIPPIG	DVVVATDAL	MIGFTGDFDS	1450
VIDCNICVTQ	TVDFSLDPTF	TIETTIVPOD	AVSRSQRRGR	TGRGRSGIYR	1500
FVTPGERPSG	MFDSSVLCEC	YDAGCAWYEL	TPAETSVRLR	AYLNTFGLPV	1550
CQDHLEFWES	VFTGLTHIDA	HFLSQTKQAG	DNEPYLVAYQ	ATVCARAQAP	1600
PPSWDQMWKC	LIRLKPTLHG	PTPLLYRLGA	VQNEVILTHP	ITKYIMACMS	1650
ADLEVVTSTW	VLNGGVLAAL	AAYCLTTGSV	VTVGRIILSG	KPAVVPDREV	1700
LYQEFDEMEE	CASQLPYIEQ	GMQLAEQFKQ	KALGLLQTAT	KQAEAAAPVV	1750
ESKWRALETF	WAKHMANFIS	GIQYLAGLST	LPGNPAIASL	MAFTASITSP	1800
LTTQNTLLFN	ILGGWVAAQL	APPSAASAFV	GAGIAGA AVG	SIGLGKVLVD	1850
ILAGYGAGVA	GALVAFKVM	GEVPSTEDLV	NLLPAILSPG	ALVVGVCVCA	1900

FIG. 14G

SUBSTITUTE SHEET (RULE 26)

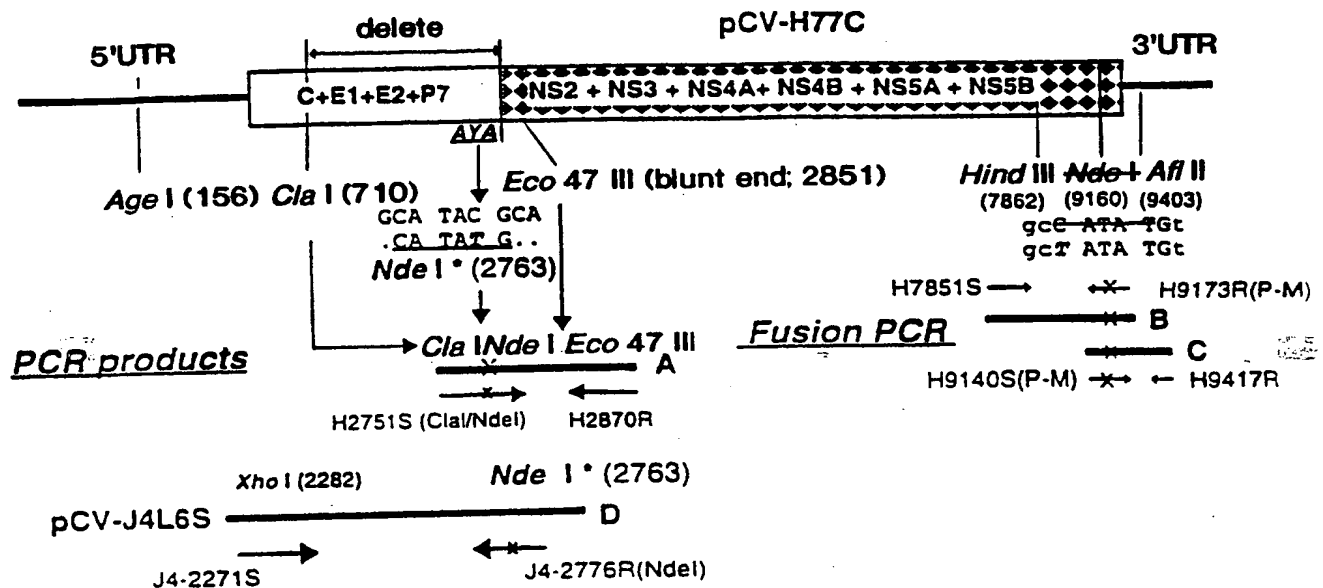
31/49

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ILRRHVGPGE	GAVQAMNRLI	AFASRGNHVS	PIHYVPESDA	AARVTQILSS	1950
LTTTQLLKRL	HQWINECST	PCSGSWLRDV	WDWICTVLTD	FKTWLQSKLL	2000
PRLPGVPFLS	CQRGYKGVWR	GCGIMQTTCP	CGAQIAGHVK	NGSMRIVGPR	2050
TCSNIWHGTF	PINAYTTGPC	TPSPAIFYSR	ALWRVAAEEY	VEVTRVGDFH	2100
YVTGMTIDNV	KCPCQVPAPE	FFTEVDGVRL	HRYPACKPL	LREDVTFQVG	2150
LNQYLVGSOL	PCEPEPDVIV	LTSMLTDPST	ITAETAKRRL	ARGSPPSLAS	2200
SSASQLSAPS	LKATCTIHD	SPDADLIFAN	LLWRQEMGGN	IIRVESENKV	2250
VILDSFEPLH	AEGDEREISV	AAEILRKSRK	FPSALPIWAR	PDYNPELLES	2300
WKDPDYVPPV	VHGCPLEPTK	APPITPPRRK	RTVVLTESNV	SSALAEIATK	2350
TFGSSGSSAV	DSGTATALPD	LASDDGDKGS	DVESYSSMPP	LEGEFGDPDL	2400
SDGSWSIVSE	EASEDVCCS	MSYTWIGALI	TPCAAEEKSL	PINPLSNLL	2450
RHHNMVYATT	SRSASLRQKK	VIFDRLQVLD	DHYRDLKEM	KAKASTVKAK	2500
LLSIEEACKL	TPPHSAKSKF	GYGAKDVRNL	SSRAVNHRS	WEDLLEDTE	2550
TPIDTTIMAK	SEVFCVQPEK	GGRKPARLIV	FDDLGVVCE	KMALYDVVST	2600
LPQAVMGSSY	GFQYSPKQRV	EFLVNTWWSK	KCPMGFSYDT	RCFDSTVTE	2650
DIRVEESTYQ	CCDLAPEARQ	AIRSLTERLY	IGGPLTNSKG	QNGYRRRCRA	2700
SGVLTTSCGN	TLTCYLKATA	ACRAAKLQDC	TMLVNGDDL	VICESAGIQE	2750
DAAALRAFTE	AMTRYSAFPG	DPPQPEYDLE	LITSCSSNVS	VAHDASGKRV	2800
YYLTRDPTTP	LARAAWETAR	HTPINSWLG	IIMYAPTLWA	RMILMIHFFS	2850
ILLAQEULEK	ALDCQTYGAC	YSIEPLDLPO	IIEERLHGLSA	FTLHSYSPGE	2900
INRVASCLRK	LGVPLRIWR	HRARSVRACL	LSQGGRAATC	GRYLFNWAVR	2950
TKLKLTPIPA	ASQLDLGWF	VAGYSGGDIY	HSLSRARPRW	FPLCLLLLSV	3000
GVGIYLLPNR					3010

FIG. 14H

32/49

**#2. Strategy for constructing chimeric clone of HCV (pH77CV-J4) which contains the nonstructural region of strain H77 and the structural region of strain HC-J4**



1. Fragment A, B, C and D ; PCR amplification from pCV-H77C or pCV-J4L6S
  - Fragment A ; additional *Cla* I site, artificial *Nde* I site induced by a single mutation (C→T at nt 2765 of H77C) and authentic *Eco*47 III site
  - Fragment B and C ; eliminated *Nde* I site by a single mutation within the primers (C→T at nt 9158 of H77C) , and fusion PCR with both fragments
  - Fragment D ; artificial *Nde* I site induced by 2 point mutations within the primer (T→A at nt 2762 and C→T at nt 2765 of J4L6S)
2. TA cloning of PCR products
3. Sequence analysis
4. Cloning of Fragment A (*Cla* I-*Eco* 47III ) and Fragment B/C (*Hind* III-*Afl* II ) with correct sequence into pCV-H77C
5. Complete sequence analysis of new cassette vector [pH77CV], into which the structural regions of different genotypes can be inserted.
6. Cloning of Fragment-Age I/*Xho* I (cut out from pCV-J4L6S) and Fragment D (*Xho* I-*Nde* I) with correct sequence into the new cassette vector ; 3 piece ligation
7. Complete sequence analysis of 1a+1b chimera [pH77CV-J4]
8. *In vitro* transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 15

33/49

GCCAGCCCCC	TGATGGGGGC	GACACTCCAC	CATGAATCAC	TCCCCGTGTA	50
GGAACACTG	TCTTCACGCA	GAAAGCGTCT	AGCCATGGCG	TTAGTATGAG	100
TGTGGTGCAG	CCTCCAGGAC	CCCCCCTCCC	GGGAGAGCCA	TAGTGGTCTG	150
CGGAACCGGT	GAGTACACCG	GAATTGCCAG	GACGACCGGG	TCTTTTCTTG	200
GATCAACCCG	CTCAATGCCT	GGAGATTITG	GCGTGGCCCC	GCGAGACTGC	250
TAGCCGAGTA	GTGTTGGGTC	GCGAAAGGCC	TTGTGGTACT	GCCTGATAGG	300
GTGCTTGCGA	GTGCCCCGGG	AGGTCTCGTA	GACCGTGCAC	CATGAGCAAG	350
AATCCTAAAC	CTCAAAGAAA	AACCAAAAGT	AACACCAACC	GCGGCCCCCA	400
GGACGTCAAG	TTCCCCGGCG	GTGGTCAAGT	CGTTGGTGGG	GTTTACCTGT	450
TGCCCCGCGC	GGGCCCCAGG	TTGGGTGTGC	GCGCGACTAG	GAAGGCTTCC	500
GAGCGGTCCG	AACCTCGTGG	AAGGCGACAA	CCTATCCCAA	AGGCTCGCCG	550
ACCCGAGGGC	AGGGCCTGGG	CTCAGCCCCG	GTACCCCTTG	CCCCCTCTATG	600
GCAATGAGGG	CCTGGGGTGG	GCAGGATGGC	TCTGTTCACC	CCGCGGCTCC	650
CGGCGTAGTT	GGGCCCCCAC	GGACCCCCCG	CGTAGGTGGC	GTAACCTTGGG	700
TAAGGTTCATC	GATACCCCTTA	CATGCGGCTT	CGCCGATCTC	ATGGGGTACA	750
TTCCGCTCGT	CGCGCCCCCC	CTAGGGGGCG	CTGCCAGGGC	CTTGGCACAC	800
GGTGTCCGGG	TTCTGGAGGA	CGGCGTGAAC	TATGCAACAG	GGAACCTGCC	850
CGGTGCTCT	TTCTCTATCT	TCTCTTTGGC	TCTGCTGTCC	TGTTTGACCA	900
TCCCAGCTTC	CGCTTATGAA	GTGCGCAACG	TGTCCGGGAT	ATACCATGTC	950
ACGAACGACT	GCTCCAACCTC	AAGCATTTGT	TATGAGGCAG	CGGACGTGAT	1000
CATGCATACT	CCCGGGTGGC	TGCCCCGTGT	TCAGGAGGGT	AACAGCTCCC	1050
GTGTGCTGGT	AGCGCTCACT	CCCACGCTCG	CGGCCAGGAA	TGCCAGCGTC	1100
CCCACTACGA	CAATACGACG	CCACGTTCGAC	TTGCTCGTTG	GGACGCGTGC	1150
TTTCTGCTCC	GCTATGTACG	TGGGGGATCT	CTGCGGATCT	ATTTTCTCTG	1200
TCTCCCAGCT	GTTCACCTTC	TGGCCTCGCC	GCCATGAGAC	AGTGCAGGAC	1250
TGCAACTGCT	CAATCTATCC	CGGCCATGTA	TCAGGTACCC	GCATGGCTTG	1300
GGATATGATG	ATGAACCTGGT	CACCTACAAC	AGCCCTAGTG	GTGTCCGAGT	1350
TGCTCCGGAT	CCCAACAAGCT	GTGCTGGACA	TGGTGGCGGG	GGCCCACTGG	1400
GGAGTCCCTG	CGGGCCTTGC	CTACTATTCC	ATGGTAGGGA	ACTGGGCTAA	1450
GGTTCTGATT	GTGGCGCTAC	TCTTTTGGCG	CGTTGACGGG	GAGACCCACA	1500
CGACGGGGAG	GGTGGCGGGC	CACACCACT	CCGGGTTCAC	GTCCCTTTTC	1550
TCATCTGGGG	CGTCTCAGAA	AATCCAGCTT	GTGAATACCA	ACGGCAGCTG	1600
GCACATCAAC	AGGACTGCCC	TAAATTGCAA	TGACTCCCTC	CAAACTGGGT	1650
TCTTTGGCGC	GCTGTTTTAC	GCACACAAGT	TCAACTCGTC	CGGGTGGCCG	1700
GAGCGCATGG	CCAGCTGCCC	CCCCATGAC	TGGTTGCCCC	AGGGGTGGGG	1750
CCCCATCACC	TATACTAAGC	CTAACAGCTC	GGATCAGAGG	CCTTATTGCT	1800

FIG. 16A

SUBSTITUTE SHEET (RULE 26)

34/49

GGCATTACGC	GCCTCGACCG	TGTGGTGTGCG	TACCCGCGTC	GCAGGTGTGT	1850
GGTCCAGTGT	ATTGTTTCAC	CCCAAGCCCT	GTGTGTGGTGG	GGACCAACGA	1900
TGCTTCCGGT	GTCCCTACGT	ATAGCTGGGG	GGAGAATGAG	ACAGACGTGA	1950
TGCTCTCAA	CAACACGGT	CCGCCACAAG	GCAACTGGTT	CGGCTGTACA	2000
TGGATGAATA	GTACTGGGTT	CACTAAGACG	TGCGGAGGTC	CCCCGTGTAA	2050
CATCGGGGGG	GTGGGTAAAC	GCACCTTGAT	CTGCCCCACG	GACTGCTTCC	2100
GGAAGCACCC	CGAGGCTACT	TACACAAAAT	GTGGCTCGGG	GCCCTGGTTG	2150
ACACCTAGGT	GCCTAGTAGA	CTACCCATAC	AGGCTTTGGC	ACTACCCCTG	2200
CACCTCTCAAT	TTTTCATCT	TTAAGGTTAG	GATGTATGTG	GGGGGCGTGG	2250
AGCACAGGCT	CAATGCCGCA	TGCAATTGGA	CTCGAGGAGA	GCGCTGTAAAC	2300
TTGGAGGACA	GGGATAGGTC	AGAACTCAGC	CCGCTGCTGC	TGTCTACAAC	2350
AGAGTGCCAG	ATACTGCCCT	GTGCTTTCAC	CACCCTACCG	GCTTTATCCA	2400
CTGGTTTGAT	CCATCTCCAT	CAGAACATCG	TGGACGTGCA	ATACCTGTAC	2450
GGTGTAGGGT	CAGCGTTTGT	CTCCTTTGCA	ATCAAATGGG	AGTACATCCT	2500
GTTCCTTTTC	CTTCTCCTGG	CAGACGCGCG	CGTGTGTGCC	TGCTTGTGGA	2550
TGATGCTGCT	GATAGCCCGAG	GCTGAGCCCG	CCTTAGAGAA	CTTGGTGGTC	2600
CTCAATGCGG	CGTCCGTGGC	CGGAGCCCAT	GGTATTCCTCT	CCTTTCCTTGT	2650
GTTCCTCTGC	GCCGCCCTGGT	ACATTAAAGG	CAGGCTGGCT	CCTGGGCGCG	2700
CGTATGCTTT	TTATGGCGTA	TGGCCGCTGC	TCCTGCTCCT	ACTGGCGTTA	2750
CCACCACGAG	CATATGCACT	GGACACGGAG	GTGGCCGCGT	CGTGTGCCGG	2800
CGTTGTCTTT	GTCCGGTTAA	TGGCGCTGAC	TCTGTGCCCA	TATTACAAGC	2850
GCTATATCAG	CTGGTGCATG	TGGTGGCTTC	AGTATTTTCT	GACCAGAGTA	2900
GAAGCGCAAC	TGCACGTGTG	GGTTCCCCCC	CTCAACGTCC	GGGGGGGGCG	2950
CGATGCCGTC	ATCTTACTCA	TGTGTGTAGT	ACACCCGACC	CTGGTATTTG	3000
ACATCACCAA	ACTACTCCTG	GCCATCTTCG	GACCCCTTTG	GATTCTTCAA	3050
GCCAGTTTGC	TTAAAGTCCC	CTACTTCGTG	CGGTTTCAAG	GCCTTCTCCG	3100
GATCTGCGCG	CTAGCGCGGA	AGATAGCCCG	AGGTCATTAC	GTGCAAATGG	3150
CCATCATCAA	GTTAGGGGCG	CTTACTGGCA	CCTATGTGTA	TAACCATCTC	3200
ACCCCTCTTC	GAGACTGGGC	GCACAACGGC	CTGCGAGATC	TGGCCGTGGC	3250
TGTGGAACCA	GTGCTCTTCT	CCCGAATGGA	GACCAAGCTC	ATCACGTGGG	3300
GGGCAGATAC	CGCCGCGTGC	GGTGACATCA	TCAACGGCTT	GCCCGTCTCT	3350
GCCCGTAGGG	GCCAGGAGAT	ACTGCTTGGG	CCAGCCGACG	GAATGGTCTC	3400
CAAGCGGTGG	AGGTTGCTGG	CGCCCATCAC	GGCGTACGCC	CAGCAGACGA	3450
GAGGCTCTCT	AGGGTGTATA	ATCACCGACC	TGACTGGCCG	GCACAAAAC	3500
CAAGTGGAGG	GTGAGGTCCA	GATCGTGTCA	ACTGCTACCC	AAACCTTCTT	3550
GGCAACGTGC	ATCAATGGGG	TATGCTGGAC	TGTCTAACCA	GGGGCCGGAA	3600

FIG. 16B

35/49

CGAGGACCAT	CGCATCACCC	AAGGGTCCTG	TCATCCAGAT	GTATACCAAT	3650
GTGGACCAAG	ACCTTGTGGG	CTGGCCCCGT	CCTCAAGGTT	CCCGCTCATT	3700
GACACCCCTGT	ACCTGCGGCT	CCTCGGACCT	TTACCTGGTC	ACGAGGCACG	3750
CCGATGTTCAT	TCCCCGTGCG	CGGCGAGGTG	ATAGCAGGGG	TAGCCTGCTT	3800
TCGCCCCCGG	CCATTTCCTA	CTTGAAAGGC	TCCTCGGGGG	GTCCGCTGTT	3850
GTGCCCCGGG	GGACACGCGG	TGGGCTTATT	CAGGGCCGGG	GTGTGCACCC	3900
GTGGAGTGGC	TAAAGCGGTG	GACTTTTATC	CTGTGCAGAA	CCTAGGGACA	3950
ACCATGAGAT	CCCCGGTGT	CACGGACAAC	TCCTCTCCAC	CAGCAGTGCC	4000
CCAGAGCTTC	CAGGTGGCCC	ACCTGCATGC	TCCCACCGGC	AGCGGTAAAG	4050
GCACCAAGGT	CCCGGCTGCG	TACGCAGCCC	AGGGCTACAA	GGTGTGTGGT	4100
CTCAACCCCT	CTGTGTGCTG	AACGCTGGGC	TTTGGTGTCT	ACATGTCCAA	4150
GGCCCATGGG	GTGTATCCTA	ATATCAGGAC	CGGGGTGAGA	ACAATTACCA	4200
CTGGCAGCCC	CATCACGTAC	TCCACCTACG	GCAAGTTCCT	TGCGGACGGC	4250
GGGTGCTCAG	GAGGTGCTTA	TGACATAATA	ATTGTGTACG	AGTGCCACTC	4300
CACGGATGCC	ACATCCATCT	TGGGCATCGG	CAGTGTCTCT	GACCAAGCAG	4350
AGACTGCGGG	GGCGAGACTG	GTGTGTGCTG	CCACTGCTAC	CCCTCCGGGC	4400
TCCGTCACTG	TGTCCCATCC	TAACATCGAG	GAGGTGTCTC	TGTCCACCAC	4450
CGGAGAGATC	CCCTTTTACG	GCAAGGCTAT	CCCCCTCGAG	GTGATCAAGG	4500
GGGGAAGACA	TCTCATCTTC	TGCCACTCAA	AGAAGAAGTG	CGACGAGCTC	4550
GCCGCGAAGC	TGGTCCGATT	GGGCATCAAT	GCCGTGGCCT	ACTACCGGGG	4600
TCTTGACGTG	TCTGTTCATC	CGACCAGCGG	CGATGTTGTC	GTGCTGTCCA	4650
CCGATGCTCT	CATGACTGGC	TTTACCGGGG	ACTTCGACTC	TGTGATAGAC	4700
TGCAACACGT	GTGTCACTCA	GACAGTCCAT	TTTACGCTTG	ACCTTACCTT	4750
TACCATTTAG	ACAACCACGC	TCCCCCAGGA	TGCTGTCTCC	AGGACTCAAC	4800
GCCGGGGCAG	GACTGGCAGG	GGGAAGCCAG	GCATCTATAG	ATTTGTGGCA	4850
CCGGGGGAGC	GCCCCCTCGG	CATGTTTCGAC	TGGTCCGTCC	TCTGTGTAGT	4900
CTATGACCGG	GGCTGTGCTT	GGTATGAGCT	CACGCCCCGG	GAGACTACAG	4950
TTAGGCTACG	AGGTATACAT	AACACCCCGG	GGCTTCCCGT	GTGCCAGGAC	5000
CATCTTGAAT	TTTGGGAGGG	CGTCTTTACG	GGCTTCACTC	ATATAGATGC	5050
CCACTTTTAA	TCCCAGACAA	AGCAGAGTGG	GGAGAACTTT	CCCTTACCTGG	5100
TAGGTATCCA	AGCCACCGTG	TGGGCTTAGG	CTCAAGCCCC	TCCCCCATCG	5150
TGGGACCAGA	TGTGGAAGTG	TTTGATCCGC	CTTAAACCCA	CCCTCCATGG	5200
GCCAACACCC	CTGCTATACA	GACTGGGGGC	TGTTTCAGAAT	GAAGTCACCC	5250
TGACGCACCC	AATCACCAAA	TACATCATGA	CATGCATGTC	GGCCGACCTG	5300
GAGGTGCTCA	CGAGCACCTG	GGTGTCTGTT	GGCGGGGTCC	TGGCTGCTCT	5350
GGCCGCGTAT	TGCCTGTCAA	CAGGCTGGGT	GGTCATAGTG	GGCAGGATCG	5400

FIG. 16C

SUBSTITUTE SHEET (RULE 26)



36/49

TCTTGTCOOG	GAAGCCGGCA	ATTATACCTG	ACAGGGAGGT	TCTCTACCAG	5450
GAGTTCGATG	AGATGGAAGA	GTCCTCTCAG	CACCTTACCGT	ACATCGAGCA	5500
AGGGATGATG	CTCGCTGAGC	AGTTCAAGCA	GAAGGCCCTC	GGCCTCCTGC	5550
AGACCGOGTC	CCGCCATGCA	GAGGTTATCA	CCCCTGCTGT	CCAGACCAAC	5600
TGGCAGAAAC	TGGAGGTCTT	TTGGGGGAAG	CACATGTGGA	ATTTCATCAG	5650
TGGGATACAA	TACTTGGGGG	GCTGTGCAAC	GCTGCCCTGGT	AACCCCGCCA	5700
TTGCTTCATT	GATGGCTTTT	ACAGCTGCGG	TCACCAGCCC	ACTAACCCTT	5750
GGCCAAACCC	TCCTCTTCAA	CATATTGGGG	GGGTGGGTGG	CTGCCACGCT	5800
CGCCGCCCCC	GGTCCCGCTA	CTGCCCTTGT	GGGTGCTGGC	CTAGCTGGGG	5850
CCGCCATCGG	CAGCGTTGGA	CTGGGGGAAG	TCCTCGTGGG	CATTCTTGCA	5900
GGGTATGGCG	CGGGCGTGGC	GGGAGCTCTT	GTAGCATTCA	AGATCATGAG	5950
CGGTGAGGTC	CCCTCCACGG	AGGACCTGGT	CAATCTGCTG	CCCGCCATCC	6000
TCTCGCCTGG	AGCCCTTGTA	GTCGGTGTGG	TCTGCGCAGC	AATACTGGCG	6050
CGGCACGTTG	GCCCGGGCGA	GGGGGCAGTG	CAATGGATGA	ACCGGCTAAT	6100
AGCCTTCGCC	TCCCGGGGGA	ACCATGTTTC	CCCCACGCAC	TACGTGCCGG	6150
AGAGCGATGC	AGCCGCCCGC	GTCACCTGCA	TACTCAGCAG	CCTCACTGTA	6200
ACCCAGCTCC	TGAGGCGACT	GCATCAGTGG	ATAAGCTCGG	AGTGTACCAC	6250
TCCATGCTCC	GGTTCCTGGC	TAAGGGACAT	CTGGGACTGG	ATATGCGAGG	6300
TGCTGAGCGA	CTTTAAGACC	TGGCTGAAAG	CCAAGCTCAT	GCCACAACCTG	6350
CCTGGGATTTC	CCTTTGTGTC	CTGCCAGCGC	GGGTATAGGG	GGGTCTGGCG	6400
AGGAGACGGC	ATTATGCCACA	CTCGCTGCCA	CTGTGGAGCT	GAGATCACTG	6450
GACATGTCAA	AAACGGGACG	ATGAGGATCG	TGGTTCCTAG	GACCTGCAGG	6500
AACATGTGGA	GTTGGACGTT	CCCCATTAAAC	GCCTACACCA	CGGGCCCTTG	6550
TACTCCCCCT	CCTGCGCCGA	ACTATAAGTT	CGCGCTGTGG	AGGGTGTCTG	6600
CAGAGGAATA	CGTGGAGATA	AGGCGGGTGG	GGGACTTCCA	CTACGTATCG	6650
GGTATGACTA	CTGACAATCT	TAAATGCCCG	TGCCAGATCC	CATCGCCCGA	6700
ATTTTTCACA	GAATTGGACG	GGGTGCGCCT	ACACAGGTTT	GCGCCCCCTT	6750
GCAAGCCCTT	GCTGCGGGAG	GAGGTATCAT	TCAGAGTAGG	ACTCCACGAG	6800
TACCCGGTGG	GGTCCGAATT	ACCTTGGCAG	CCCGAACCGG	ACGTAGCCGT	6850
GTTGACGTCC	ATGCTCACTG	ATCCCTCCCA	TATAACAGCA	GAGGCGGCGG	6900
GGAGAAGGTT	GGCGAGAGGG	TCACCCCTTT	CTATGGCCAG	CTCCTCGGCT	6950
AGCCAGCTGT	CCGCTCCATC	TCTCAAGGCA	ACTTGCACCG	CCAACCATGA	7000
CTCCCCCTGAC	GCCGAGCTCA	TAGAGGCTAA	CCTCCTGTGG	AGGCAGGAGA	7050
TGGGCGGCAA	CATCACCAGG	GTTGAGTCAG	AGAACAAGT	GGTGATTCTG	7100
GACTCCTTCG	ATCCGCTTGT	GGCAGAGGAG	GATGAGCGGG	AGGTCTCCGT	7150
ACCTGCAGAA	ATTCTGCGGA	AGTCTCGGAG	ATTGCGCCCG	GCCCTGCCCG	7200

FIG. 16D

37/49

TCTGGGGCGCG	GCCGGACTAC	AACCCCCCGC	TAGTAGAGAC	GTGGAAAAG	7250
CCTGACTACG	AACCACCTGT	GGTCCATGGC	TGCCCCGTAC	CACCTCCAAG	7300
GTCCCCCTCT	GTGCTCTCGC	CTCGGAAAAA	GCGTACGGTG	GTCTTCACCG	7350
AATCAACCCCT	ATCTACTGCC	TTGGCCGAGC	TTGCCACCAA	AAGTTTTTGGC	7400
AGCTCCTCAA	CTTCCGGCAT	TACGGGGGAC	AATACGACAA	CATCCTCTGA	7450
GCCCCCCCCCT	TCTGGCTGCC	CCCCCGACTC	CGACGTTGAG	TCCATATCTT	7500
CCATGCCCCC	CCTGGAGGGG	GAGCCTGGGG	ATCCGGATCT	CAGCGACGGG	7550
TCATGGTCCA	CGGTACGTAG	TGGGGCCGAC	ACGGAAGATG	TGGTGTGCTG	7600
CTCAATGTCT	TATTCCTGGA	CAGGGGCACT	CGTCACCCCG	TGGGCTGGGG	7650
AAGAACAAAA	ACTGCCCCATC	AACGCACTGA	GCAACTCGTT	GCTACGGCCAT	7700
CACAATCTGG	TGTATTCCAC	CACTTCACGC	AGTGCTTGCC	AAAGGCAGAA	7750
GAAAGTCACA	TTTGACAGAC	TGCAAGTTCT	GGACAGCCAT	TACCAGGACG	7800
TGCTCAAGGA	GGTCAAAGCA	GCGGCGTCAA	AAGTGAAGGC	TAACCTTGCTA	7850
TCCGTAGAGG	AAGCTTGCAG	CCTGACGCCC	CCACATTTCAG	CCAAATCCAA	7900
GTTTGGCTAT	GGGGCAAAAG	ACGTCCGTTG	CCATGCCAGA	AAGGCCGTAG	7950
CCCACATCAA	CTCCGTGTGG	AAAGACCTTC	TGGAAGACAG	TGTAACACCA	8000
ATAGACACTA	CCATCATGGC	CAAGAACGAG	GTTTTCTGCG	TTCAGCCTGA	8050
GAAGGGGGGT	CGTAAGCCAG	CTCGTCTCAT	CGTGTTCCCC	GACCTGGGGG	8100
TGCGCGTGTG	CGAGAAGATG	GCCCTGTACG	ACGTGGTTCAG	CAAGCTCCCC	8150
CTGGCCGTGA	TGGGAAGCTC	CTACGGATTTC	CAATACTCAC	CAGGACAGCG	8200
GGTTGAATTC	CTCGTGCAAG	CGTGGGAAGTC	CAAGAAGACC	CCGATGGGGT	8250
TCTCGTATGA	TACCCGCTGT	TTTGACTCCA	CAGTCACTGA	GAGCGACATC	8300
CGTACGGAGG	AGGCAATTTA	CCAATGTTGT	GACCTGGACC	CCCAAGCCCC	8350
CGTGGCCATC	AAGTCCCTCA	CTGAGAGGCT	TTATGTTGGG	GGCCCTCTTA	8400
CCAATTCAAG	GGGGGAAAC	TGCGGCTACC	GCAGGTGCCC	CGCGAGCGGC	8450
GTACTGACAA	CTAGCTGTGG	TAACACCCCTC	ACTTGCTACA	TCAAGGCCCC	8500
GGCAGCCTGT	CGAGCCGCAG	GGCTCCAGGA	CTGCACCATG	CTCGTGTGTG	8550
GCGACGACTT	AGTCGTTATC	TGTGAAGATG	CGGGGGTCCA	GGAGGACGGG	8600
GCGAGCCTGA	GAGCCTTCAC	GGAGGCTATG	ACCAGGTACT	CCGCCCCCCC	8650
CGGGGACCCC	CCACAACCAG	AATACGACTT	GGAGCTTATA	ACATCATGCT	8700
CCTCCAACGT	GTCAGTCGCC	CACGACGGCG	CTGGAAAGAG	GGTCTACTAC	8750
CTTACCCGTG	ACCTTACAAC	CCCCCTCGCG	AGAGCCCGGT	GGGAGACAGC	8800
AAGACACACT	CCAGTCAATT	CCTGGCTAGG	CAACATAATC	ATGTTTGCCC	8850
CCCACTGTGT	GGCGAGGATG	ATACTGATGA	CCCATTTCTT	TAGCGTCTTC	8900
ATAGCCAGGG	ATCAGCTTGA	ACAGGCTCTT	AACTGTGAGA	TCTACGGAGC	8950
CTGCTACTCC	ATAGAACCAC	TGGATCTACC	TCCAATCATT	CAAAGACTCC	9000

FIG. 16E

38/49

ATGGCCTCAG	CGCATTTTCA	CTCCACAGTT	ACTCTCCAGG	TGAAATCAAT	9050
AGGGTGGCCG	CATGCCTCAG	AAAACTTGGG	GTCCCGCCCT	TGCGAGCTTG	9100
GAGACACCGG	GCCCCGAGCG	TCCCGGCTAG	GCTTCTGTCC	AGAGGAGGCA	9150
GGGCTGCTAT	ATGTGGCAAG	TACCTCTTCA	ACTGGGCAGT	AAGAACAAG	9200
CTCAAACCTCA	CTCCAATAGC	GGCCGCTGGC	CGGCTGGACT	TGTCCGGTTG	9250
GTTCACGGCT	GGCTACAGCG	GGCGAGACAT	TTATCACAGC	GTGTCTCATG	9300
CCCGGCCCCG	CTGGTCTTGG	TTTTCGCTAC	TCCTGCTCGC	TGCAGGGGTA	9350
GGCATCTACC	TCCTCCCCAA	CCGATGAAGG	TTGGGGTAAA	CACTCCGGCC	9400
TCTTAAGCCA	TTTCCTGTTT	TTTTTTTTTT	TTTTTTTTTT	TTTTTCTTTT	9450
TTTTTTTCTT	TCCTTTCCCT	CTTTTTTTCC	TTTCTTTTTT	CCTTCTTTAA	9500
TGGTGGCTCC	ATCTTAGCCC	TAGTCACGGC	TAGCTGTGAA	AGGTCCGTGA	9550
GGCGCATGAC	TGCAGAGAGT	GCTGATACTG	GCCTCTCTGC	AGATCATGT	9599

FIG. 16F

39/49

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
MSTNPKPQRK	TKRNINRRPQ	DVKFPGGGQI	VGGVYLLPRR	GPRLGVRATR	50
KASERSQPRG	RRQPIPKARR	PEGRAWAQPQ	YFWPLYGNEG	LGWAGWLLSP	100
RGSRPSWGPT	DPRRRSRNLG	KVIDILTQGF	ADLMGYIPLV	GAPLGGAARA	150
LAHGVRVLED	GVNYATGNLP	GCSFSIFLLA	LLSCLTIPAS	AYEVRNVSGI	200
YHVINDCSNS	SIVYEADVI	MHTPGCVPCV	QEGNSSRCW	ALITPILAARN	250
ASVPTTTIRR	HVDLLVGTA	FCSAMYVDL	CGSIFLVSQ	FTFSPRRHET	300
VQDCNCSTYP	GHVSGHRMAW	DMMNWSPTT	ALVVSQLLRI	PQAVVDMVAG	350
AHWGVLGLA	YYSMVGNWAK	VLTVALLFAG	VDGETHTTGR	VAGHTTSGFT	400
SLFSSGASQK	IQLVNINGSW	HINRTALN	DSLQIGFFAA	LFYAHKFESS	450
GCPERMASCR	PIDWFAQGWG	PITYIKENSS	DQRPYOWHYA	PRPCGVVPAS	500
QVCGPVYCF	PSPVAVGTID	RSQVPTYSWG	ENETDVMILL	NIRPPQGNWF	550
GCTWMNSTGF	TKTCGGPPCN	IGGVGNRTLI	CPTDCFRKHP	EATYTKCGSG	600
PWLTPRCLVD	YPYRLWHYPC	TLNFSIFKVR	MYVGGVEHRL	NAACNWIRGE	650
RCNLEDRDRS	ELSPILLSTT	EWQILPCFT	TLPALSTGLI	HLHQNTVDVQ	700
YLYGVGSFV	SFAIKWEYIL	LLFLLADAR	VCACIWMMLL	IAQAEAALEN	750
LVVLNAASVA	GAHGILSFLV	FFCAAWYIKG	RLAPGAAYAF	YGVWPLLLLL	800
LALPFRAYAL	DTEVAASCGG	VVLVGLMALT	LSPYKRYIS	WCMWMLQYFL	850
TRVEAQLHW	VPPLNVRGGR	DAVILLMCV	HPTLVFDITK	LLLAIFGPLW	900
ILQASLLKVP	YFVRVQGLLR	ICALARKIAG	GHYVQMAIK	LGALTGTIVY	950
NHLTPLRDWA	HNGLRDLAVA	VEPVVFSRME	TKLFTWGADT	AACGDIINGL	1000
FVSARRGQEI	LLGPADGMVS	KGWRLAPIT	AYAQQTRGLL	GCIITSLTGR	1050
DKNQVEGEVQ	IVSTATQTF	ATCINGVCWT	VYHGAGTRTI	ASPKGFVIQM	1100
YTNVDQDLVG	WPAPQGSRL	TPCTCGSSDL	YLVTRHADVI	FVRRRGDSRG	1150
SLLSRPISY	LKGSSGGPLL	CPAGHAVGLF	RAAVCTRGVA	KAVDFIPVEN	1200
LGTIMRSPVF	TDNSSPPAVP	QSFQVAHLHA	PTGSGKSTKV	PAAYAAQGYK	1250
VLVLNPSVAA	TLGFGAYMSK	AHGVDNIRT	GVRTTTTGSP	ITYSTYGFEL	1300
ADGGCSGGAY	DIIICDECHS	TDATSILGIG	TVLDQAETAG	ARLVVLATAT	1350
PPGSVTVSHP	NIEEVALSTT	GEIPFYGKAI	PLEVIKGRH	LIFCHSKKKC	1400
DELAACKLVAL	GINAVAYYRG	LDSVIPTSG	DVVVSTIDAL	MITGFTGDFDS	1450
VIDCNTCVIQ	TVDFSLDPTF	TIETTTLPQD	AVSRTQRRGR	TGRGKPGIYR	1500
FVAPGERPSG	MFDSSVLCEC	YDAGCAWYEL	TPAETTVRLR	AYMNTFGLPV	1550
CQDHLEFWEG	VFTGLTHIDA	HFLSQTKQSG	ENFPYLVAYQ	ATVCARAQAP	1600
PPSWDQMNKC	LIRLKPTLHG	PTPLLYRLGA	VQNEVTLTHP	ITKYIMTOMS	1650
ADLEVVTSTW	VLVGGVLAAL	AAYCLSTGCV	VTVGRIVLSG	KPAIIPDREV	1700
LYQEFDEMEE	CSQHLFYIEQ	GMLAEQFKQ	KALGLLOTAS	RHAEVTPAV	1750
QTNWQKLEVF	WAKHMANFIS	GIQYLAGLST	LPCNPATIASL	MAFTAAVTSP	1800
LTTGQTLLEN	ILGGWAAQL	AAFGAATAFV	GAGLAGAAIG	SVGLGKVLVD	1850
ILAGYGAGVA	GALVAFKIMS	GEVPSTEDLV	NLLPAILSPG	ALVVGWCAA	1900

FIG. 16G

SUBSTITUTE SHEET (RULE 26)

40/49

## H77CV-J4aa Sequence

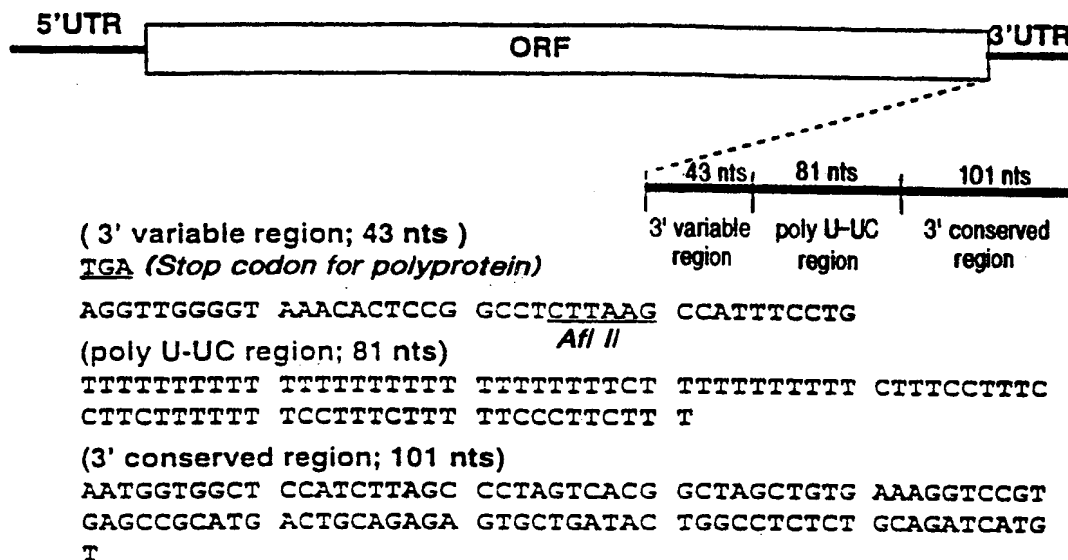
10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ILRRHVGPE	GAVQWMNRLI	AFASRGNHVS	PIHYVPESDA	AARVTAILSS	1950
LIVTQLLRRL	HQWISSECTT	PCSGSWLRDI	WDWICEVLSD	EKIWLKAKLM	2000
PQLPGIPFVS	CQRGYRGWR	GDGIMHIRCH	CGAETTGHVK	NGIMRIVGPR	2050
TCRNMWSGTF	PINAYTTGPC	TPLPAPNYKF	ALWRVSAEEY	VEIRRVGDFH	2100
YVSGMTIDNL	KCPCQIPSPE	FFTELDGVRL	HRFAPPCKPL	LREEVSFRVG	2150
LHEYFVGSQ	PCEPEPDVAV	LTSMLTDPHS	TTAEAAGRRL	ARGSPPSMAS	2200
SSASQLSAPS	LKATCTANHD	SPDAELLEAN	LLWRQEMGCN	ITRVESENKV	2250
VILDSFDPLV	AEEDEREVS	PAEILRKSRR	FARALFWAR	PDYNPPLVET	2300
WKKPDYEPFV	VHGCPLPPPR	SPPVPPPRKK	RTVVLTESTL	STALAEIATK	2350
SFGSSSTSGI	TGDNITTSSE	PAPSGCPPDS	DVESYSSMPP	LEGEFGDPDL	2400
SDGSWSTVSS	GADTEDVCC	SMSYSWIGAL	VTPCAAEEQK	LPINALSNSL	2450
LRHHNLVYST	TSRSACQROK	KVTFDRLOVL	DSHYQDVLKE	VKAAASKVKA	2500
NLLSVEEACS	LTPPHSAKSK	FGYGAKDVRC	HARKAVAHIN	SWKDLLED	2550
VTPIDTTIMA	KNEVFCVQPE	KGGRKPARLI	VFPDLGVRVC	EKMALYDVVS	2600
KLPLAVMGSS	YGFQYSPGQR	VEFLVQAWKS	KKTEMEFSYD	TRCFDSTVTE	2650
SDIRTEEATY	QCCDLDPQAR	VAIKSLTERL	YVGGPLTNSR	GENCGYRRCR	2700
ASGVLTTSCG	NILTCYIKAR	AACRAAGLQD	CTMLVCGDDL	WICESAGVQ	2750
EDAASLRAFT	EAMTRYSAAP	GDPPQPEYDL	ELITSCSSNV	SVAHDGAGKR	2800
VYYLTRDPTT	PLARAAWETA	RHTPVNSWL	NIIMFAPTLW	ARMILMIHFF	2850
SVLIARDQLE	QALNCETYGA	CYSIEPLDLP	PIIQRLHGLS	AFSLHSYSPG	2900
EINRVAACLR	KLGVPPLRAW	RHRARSVRAR	LLSRGGRAAI	CGKYLFWAV	2950
RTKLKLTPIA	AAGRDLDSGW	FTAGYSGGDI	YHSVSHARPR	WFWFCLLLLA	3000
AGVGTYLLPN	R				3011

FIG. 16H

41/49

## #1a. 3' Deletion mutants of pCV-H77C

Sequence of 3' untranslated region of pCV-H77C



## #1a -1. pCV-H77C(-98X) ; 3' 98 nucleotides removed from pCV-H77C

TGAAGGTTGG GGTAAACACT CCGGCCTCTT AAGCCATTTC CTGTTTTTTT  
 TTTTTTTTTT TTTTTTTTTT TCTTTTTTTT TTTCTTTCCT TTCCTTCTTT  
 TTTTCCTTTC TTTTCCCTT CTTTAAT

## #1a -2. pCV-H77C(-42X) ; 3' 42 nucleotides removed from pCV-H77C

TGAAGGTTGG GGTAAACACT CCGGCCTCTT AAGCCATTTC CTGTTTTTTT  
 TTTTTTTTTT TTTTTTTTTT TCTTTTTTTT TTTCTTTCCT TTCCTTCTTT  
 TTTTCCTTTC TTTTCCCTT CTTTAATGGT GGCTCCATCT TAGCCCTAGT  
 CACGGCTAGC TGTGAAAGGT CCGTGAGCCG CAT

## #1a -3. pCV-H77C(X-52) ; All of the 3' UTR sequence, except 3' 49 nucleotides, removed from pCV-H77C

TGAGCCGCAT GACTGCAGAG AGTGCTGATA CTGGCCTCTC TGCAGATCAT  
 GT

FIG. 17A

42/49

#1a -4. pCV-H77C(X) ; All of the 3' UTR sequence, except 3' 101 nucleotides, removed from pCV-H77C

TGAAATGGTG GCTCCATCTT AGCCCTAGTC ACGGCTAGCT GTGAAAGGTC  
CGTGAGCCGC ATGACTGCAG AGAGTGCTGA TACTGGCCTC TCTGCAGATC  
ATGT

#1a -5. pCV-H77C(+49X) ; The proximal 49 nucleotides of the 3' conserved region ( 98 nucleotides ; AAT not included) removed from pCV-H77C

TGAAGGTTGG GGTAACACT CCGGCCTCTT AAGCCATTTC CTGTTTTTTT  
TTTTTTTTTT TTTTTTTTTT TCTTTTTTTT TTTCTTTCCT TTCCTTCTTT  
TTTTCCTTTC TTTTCCCTT CTTTAATGCC GCATGACTGC AGAGAGTGCT  
GATACTGGCC TCTCTGCAGA TCATGT

#1a -6. pCV-H77C(VR-24) ; First 24 nucleotides of the 3' variable region removed from pCV-H77C

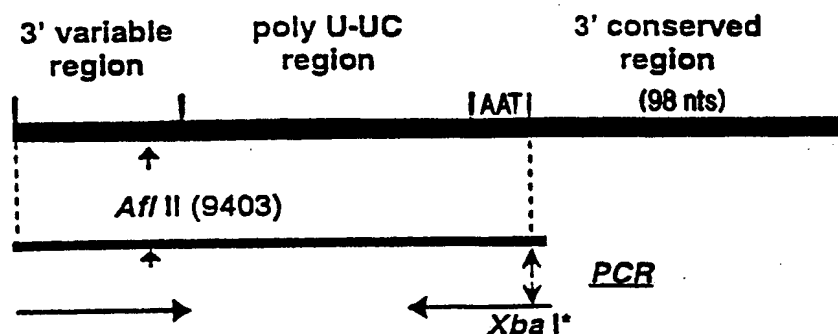
TGACTTAAGC CATTTCTGT TTTTTTTTTT TTTTTTTTTT TTTTTTCTT  
TTTTTTTTTC TTTCCTTCC TTCTTTTTTT CCTTCTTTT TCCCTTCTTT  
AATGGTGGCT CCATCTTAGC CCTAGTCACG GCTAGCTGTG AAAGGTCCGT  
GAGCCGCATG ACTGCAGAGA GTGCTGATAC TGGCCTCTCT GCAGATCATG  
T

#1a -7. pCV-H77C(-U/UC) ; Poly U-UC region removed from pCV-H77C

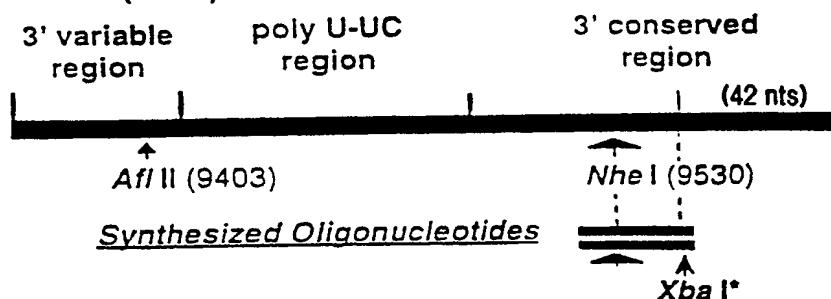
TGAAGGTTGG GGTAACACT CCGGCCTCTT AAGCCATTTC CTGAATGGTG  
GCTCCATCTT AGCCCTAGTC ACGGCTAGCT GTGAAAGGTC CGTGAGCCGC  
ATGACTGCAG AGAGTGCTGA TACTGGCCTC TCTGCAGATC ATGT

**FIG. 17B**

43/49

**#1b. Strategy of 3' Deletion mutants****#1b -1. pCV-H77C(-98X)**

1. PCR Amplification
2. Purification of PCR products.
3. Digestion with *Afl* II and *Xba* I
4. Cloning of *Afl* II /*Xba* I fragment into pCV-H77C
5. Complete sequence analysis
6. in vitro transcription (within 24 hours of inoculation)
7. Percutaneous intrahepatic transfection into chimpanzee ; 11/26/97 and 12/17/97
8. Result : Negative ( No replication)

**#1b -2. pCV-H77C(-42X)**

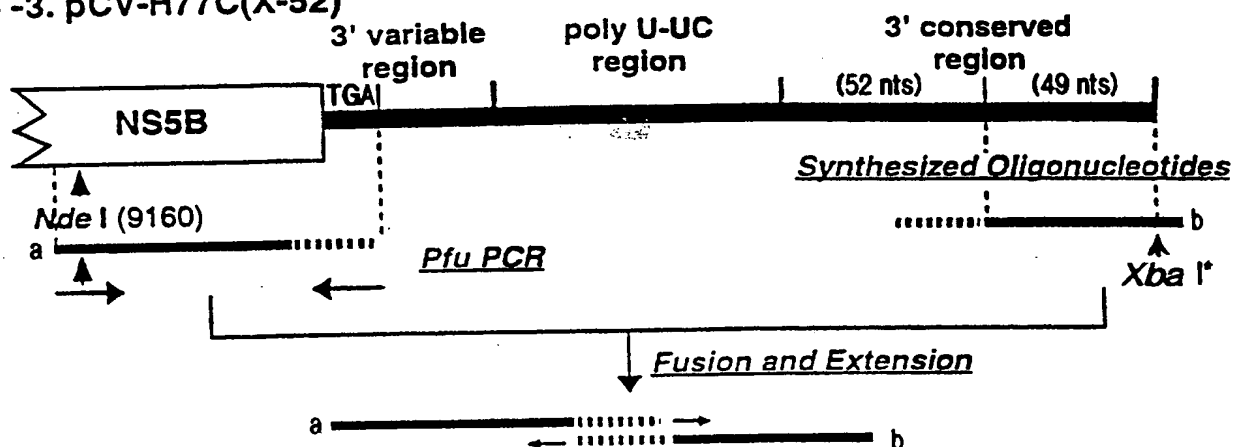
1. Synthesis of oligonucleotides ( sense and anti-sense )
2. Hybridization of oligonucleotides
3. Digestion with *Nhe* I and *Xba* I
4. Cloning of *Nhe* I /*Xba* I fragment into pG9-KL26 (3' UTR of H77C)
5. Sequence analysis
6. Cloning of 3' UTR ( -42X ) [*Afl* II /*Xba* I fragment] into pCV-H77C
7. Complete sequence analysis
8. in vitro transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee (Schedule; 1/22/98, 2/5/98 )

**FIG. 17C**



44 / 49

## #1b -3. pCV-H77C(X-52)

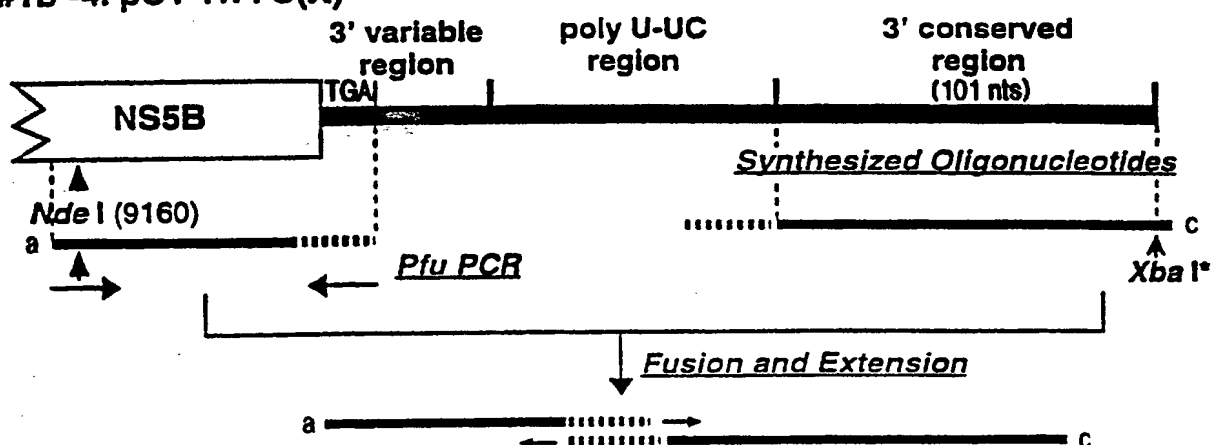


1. Fragment a ; *Pfu* PCR amplification and purification
2. Fragment b ; Synthesized oligonucleotides (anti-sense)
3. Fusion and extension
4. TA cloning
5. Sequence analysis
6. Cloning *Nde* I-*Xba* I fragment with correct sequence into pCV-H77C
7. Complete sequence analysis
8. *In vitro* transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 17D

45/49

## #1b -4. pCV-H77C(X)

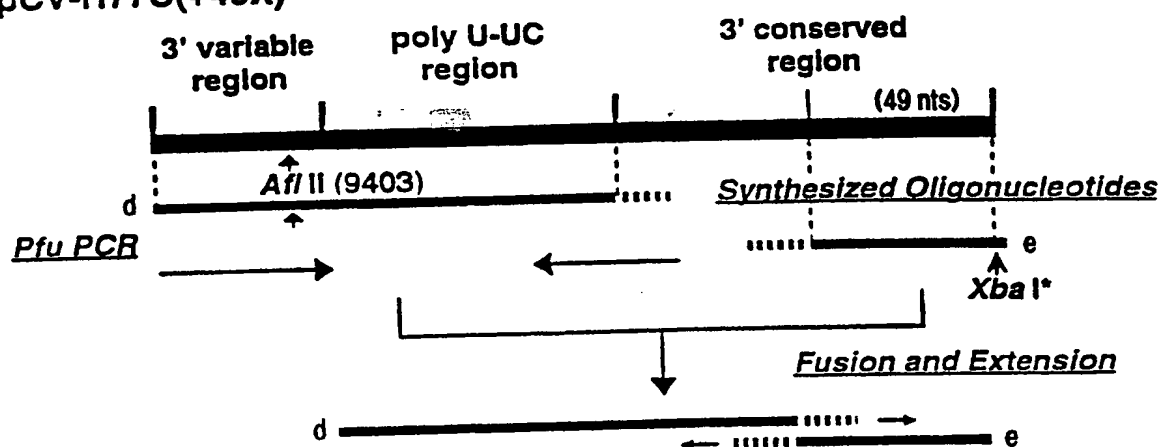


1. Fragment a ; *Pfu* PCR amplification and purification
2. Fragment c ; Synthesized oligonucleotides (anti-sense)
3. Fusion and extension
4. TA cloning
5. Sequence analysis
6. Cloning *Nde* I-*Xba* I fragment with correct sequence into pCV-H77C
7. Complete sequence analysis
8. *In vitro* transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 17E

46/49

## #1b -5. pCV-H77C(+49X)

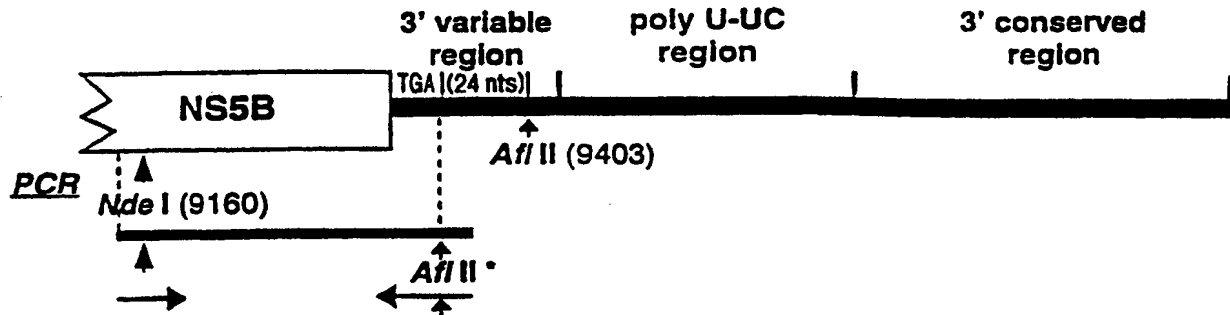


1. Fragment d ; *Pfu* PCR amplification and purification
2. Fragment e ; Synthesized oligonucleotides (anti-sense)
3. Fusion and extension
4. TA cloning
5. Sequence analysis
6. Cloning *Afl* II-*Xba* I fragment with correct sequence into pCV-H77C
7. Complete sequence analysis
8. *In vitro* transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 17F

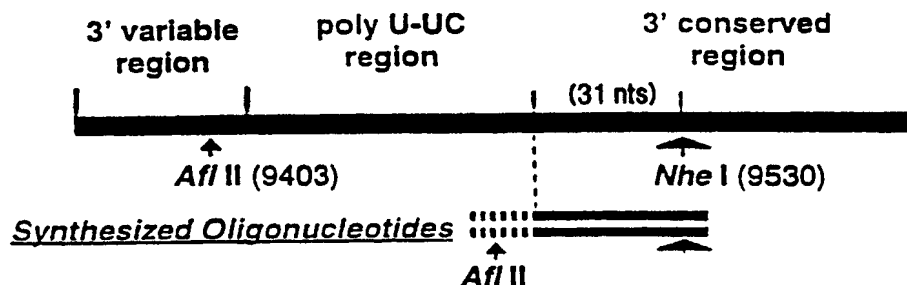
47/49

## #1b -6. pCV-H77C(VR-24)



1. PCR Amplification
2. Purification of PCR products
3. Digestion with *Nde* I and *Afl* I
4. Cloning of *Nde* I / *Afl* II fragment into pCV-H77C
5. Complete sequence analysis
6. in vitro transcription (within 24 hours of inoculation)
7. Percutaneous intrahepatic transfection into chimpanzee

## #1b -7. pCV-H77C(-U/UC)



1. Synthesis of oligonucleotides ( sense and anti-sense )
2. Hybridization of oligonucleotides
3. Digestion with *Afl* II and *Nhe* I
4. Cloning of *Afl* II and *Nhe* I fragment into pG9-KL26
5. Sequence analysis
6. Cloning of 3' UTR ( -poly U-UC ) [ *Afl* II / *Xba* I fragment ] into pCV-H77C
7. Complete sequence analysis
8. in vitro transcription (within 24 hours of inoculation)
9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 17G

48/49

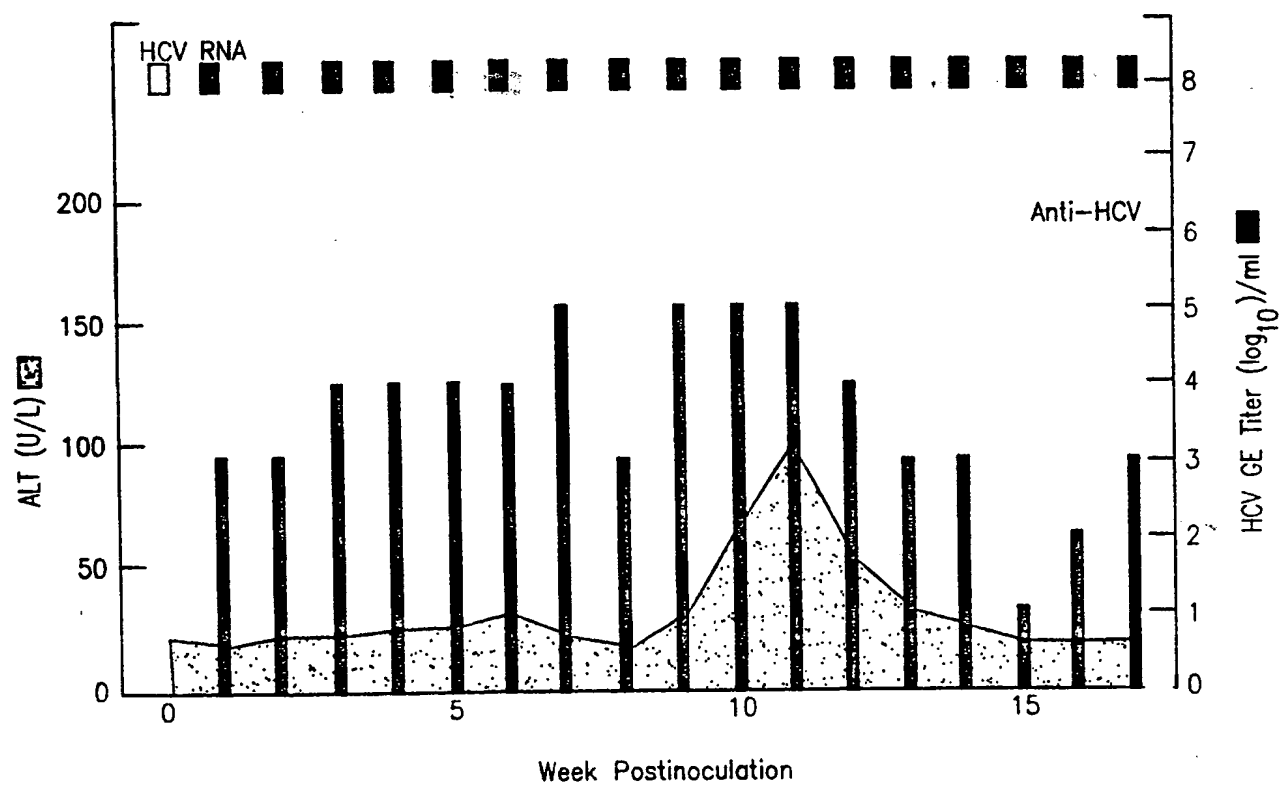


FIG. 18A

49/49

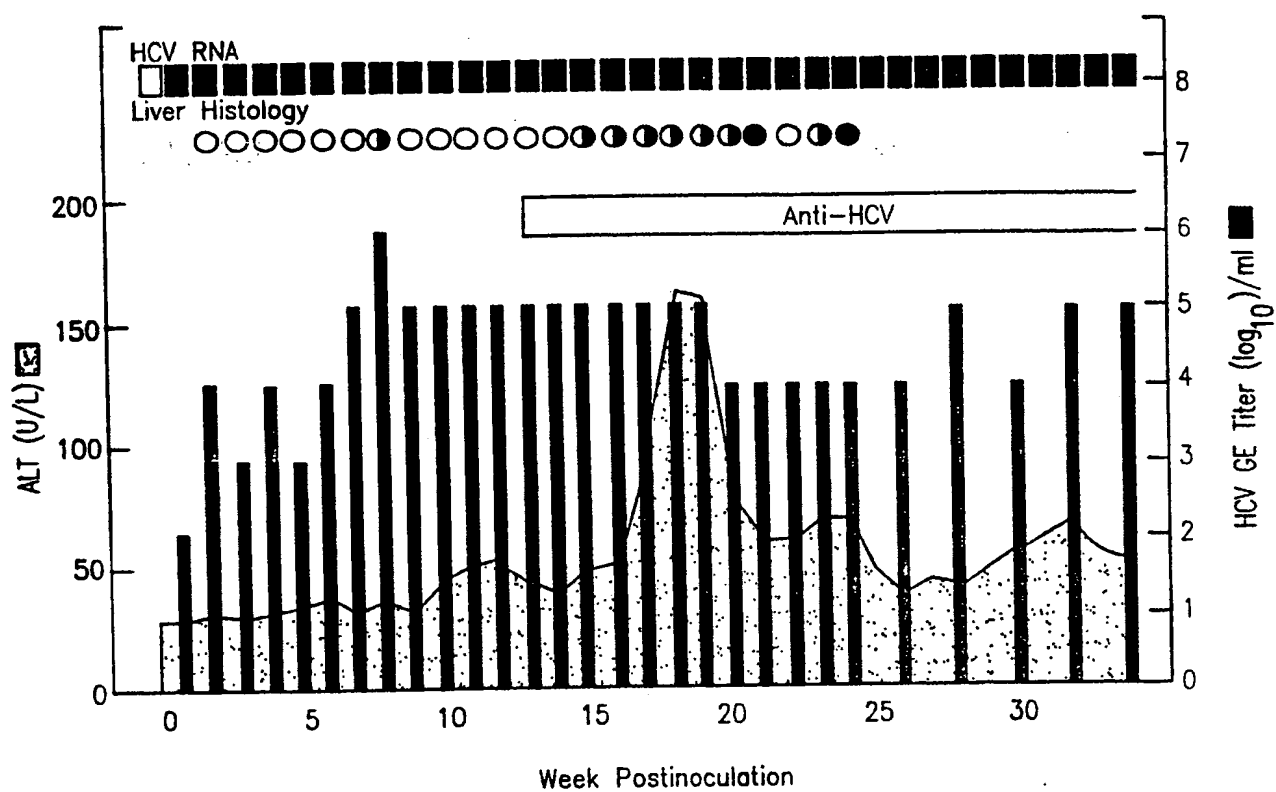


FIG. 18B



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/40, C07K 14/18, C12Q 1/70, C07K 16/10, A61K 39/29</b>		<b>A3</b>	(11) International Publication Number: <b>WO 99/04008</b>
			(43) International Publication Date: 28 January 1999 (28.01.99)
(21) International Application Number: PCT/US98/14688		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 16 July 1998 (16.07.98)		<b>Published</b> <i>With international search report.</i>	
(30) Priority Data: 60/053,062 18 July 1997 (18.07.97) US 09/014,416 27 January 1998 (27.01.98) US			
(71) Applicant: THE GOVERNMENT OF THE UNITED STATES OF AMERICA, asrepresented by THE SECRETARY, DEPARTMENT OF HEALTHAND HUMAN SERVICES [US/US]; Office of Technology Transfer, National Institutes of Health, Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).		(88) Date of publication of the international search report: 1 April 1999 (01.04.99)	
(72) Inventors: YANAGI, Masayuki; 257 <sup>th</sup> Congressional Lane, #402, Rockville, MD 20852 (US). BUKH, Jens; 7 Center Drive, MSC 0740, Bethesda, MD 20892 (US). EMERSON, Suzanne, U.; 18201 Woodcrest Drive, Rockville, MD 20852 (US). PURCELL, Robert, H.; 17517 White Grounds Road, Boyds, MD 20841 (US).			
(74) Agents: FEILER, William, S. et al.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).			
(54) Title: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND USES THEREOF			
(57) Abstract			
<p>The present invention discloses nucleic acid sequences which encode infectious hepatitis C viruses and the use of these sequences, and polypeptides encoded by all or part of these sequences, in the development of vaccines and diagnostics for HCV and in the development of screening assays for the identification of antiviral agents for HCV.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/14688

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/40 C07K14/18 C12Q1/70 C07K16/10 A61K39/29

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K C12Q A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YOO B J ET AL: "Transfection of a differentiated human hepatoma cell line (Huh7) with in vitro-transcribed hepatitis C virus (HCV) RNA and establishment of a long-term culture persistently infected with HCV" JOURNAL OF VIROLOGY, vol. 69, no. 1, January 1995, pages 32-38, XP002022696 AMERICAN SOCIETY FOR MICROBIOLOGY US see the whole document --- -/--	1,13-24, 33-35, 42,43

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

12 January 1999

Date of mailing of the international search report

22/01/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Cupido, M

# INTERNATIONAL SEARCH REPORT

Internatic Application No

PCT/US 98/14688

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	<p>WO 98 39031 A (UNIVERSITY OF WASHINGTON ;KOLYKHALOV A ; RICE C) 11 September 1998</p> <p>see SEQ ID NO:1, representing a HCV H77 consensus sequence having 99.6% identity with the nucleic acid sequences in figures 4A-4F of the present application.</p> <p style="text-align: center;">----</p>	<p>1,4,5, 13-24, 33-43</p>
X	<p>EP 0 516 270 A (JAPAN IMMUNO INC) 2 December 1992 see the whole document</p> <p style="text-align: center;">----</p>	<p>23,24, 40,41</p>
A	<p>WO 97 08310 A (WASHINGTON UNIVERSITY) 6 March 1997 see page 15; figure 3</p> <p style="text-align: center;">----</p>	<p>1-43</p>
P,X	<p>YANAGI M ET AL: "Transcripts of a chimeric cDNA clone of hepatitis C virus genotype 1b are infectious in vivo" VIROLOGY, vol. 244, no. 1, 25 April 1998, pages 161-172, XP002089701 ORLANDO US see the whole document</p> <p style="text-align: center;">-----</p>	<p>1,6-10, 13-24, 33-43</p>

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 98/14688

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
Remark: Although claim(s) 41 and 43  
is(are) directed to a method of treatment of the animal  
body, the search has been carried out and based on the alleged  
effects of the composition.
2. ☒ Claims Nos.: 29 and 30  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such  
an extent that no meaningful International Search can be carried out, specifically:  
The scope of claims 29 and 30 is so unclear and not well specified that  
a meaningful search was not possible
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all  
searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment  
of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report  
covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is  
restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US 98/14688

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9839031 A	11-09-1998	AU 6938698 A	22-09-1998
EP 0516270 A	02-12-1992	JP 5091884 A	16-04-1993
		CA 2065690 A	11-10-1992
WO 9708310 A	06-03-1997	AU 6909796 A	19-03-1997
		CA 2230452 A	06-03-1997
		EP 0856051 A	05-08-1998